National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

University of Alberta

Jackknife Methods in Robust Regression

by

Zhiyi Du

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of *Master of Science*

in

Statistics

Department of Mathematical Science

Edmonton, Alberta

Spring 1995

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

THE AUTHOR HAS GRANTED AN IRREVOCABLE NON-EXCLUSIVE LICENCE ALLOWING THE NATIONAL LIBRARY OF CANADA TO REPRODUCE, LOAN, DISTRIBUTE OR SELL COPIES OF HIS/HER THESIS BY ANY MEANS AND IN ANY FORM OR FORMAT, MAKING THIS THESIS AVAILABLE TO INTERESTED PERSONS.

L'AUTEUR A ACCORDE UNE LICENCE IRREVOCABLE ET NON EXCLUSIVE PERMETTANT A LA BIBLIOTHEQUE NATIONALE DU CANADA DE REPRODUIRE, PRETER, DISTRIBUER OU VENDRE DES COPIES DE SA THESE DE QUELQUE MANIERE ET SOUS QUELQUE FORME QUE CE SOIT POUR METTRE DES EXEMPLAIRES DE CETTE THESE A LA DISPOSITION DES PERSONNE INTERESSEES.

THE AUTHOR RETAINS OWNERSHIP OF THE COPYRIGHT IN HIS/HER THESIS. NEITHER THE THESIS NOR SUBSTANTIAL EXTRACTS FROM IT MAY BE PRINTED OR OTHERWISE REPRODUCED WITHOUT HIS/HER PERMISSION.

L'AUTEUR CONSERVE LA PROPRIETE DU DROIT D'AUTEUR QUI PROTEGE SA THESE. NI LA THESE NI DES EXTRAITS SUBSTANTIELS DE CELLE-CI NE DOIVENT ETRE IMPRIMES OU AUTREMENT REPRODUITS SANS SON AUTORISATION.

ISBN 0-612-01599-8

Canada

University of Alberta

Library Release Form

Name of Author: *Zhiyi Du*

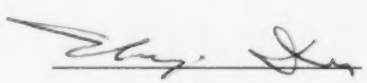
Title of Thesis: *Jackknife Methods in Robust Regression*

Degree: *Master of Science*

Year this Degree Granted: *1995*

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.



11009 89 Avenue
Edmonton, Alberta
T6G 0Z7

Date: Jan 18, 95

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Jackknife Methods in Robust Regression** submitted by **Zhiyi Du** in partial fulfillment of the requirements for the degree of **Master of Science** in **Statistics**.

Douglas P. Wiens (Supervisor)

N.G.N. Prasad

Francis Yeh

Date: Jan 10, 1995

TO MY PARENTS

ABSTRACT

Jackknife is a resample replication technique which can be used to reduce the bias of the estimator as well as to estimate the variance of the estimator. The application of jackknife procedure in the classical regression problem have been investigated since this technique was introduced by Quenouille in 1949. It has been proven that jackknife method is a useful procedure which can be used in classical regression problem when a proper type of jackknife method is applied. The objective of this thesis is to study the performance of the jackknife method in robust regression problem. A delete-one jackknife procedure is applied in obtaining the jackknife robust estimates. In consideration of the fact that the calculation of jackknife robust estimates is a time consuming procedure, a computationally efficient approximate procedure is proposed. Based on jackknife robust estimates which are discussed, a simulation study is conducted. In the later part of this thesis, several diagnostic statistics which are based on robust regression are discussed and a case study is conducted.

ACKNOWLEGEMENTS

I would like to thank all professors who have helped me during my study at the University of Alberta. In particular, I sincerely gratitude to my supervisor Dr. Douglas P. Wiens for his constant encouragement and support throughout my graduate years. This work would have been impossible without his guidance and help.

A special thanks goes to Dr. N.G.N. Prasad for his helpful suggestions and comments during the time I was working on my thesis.

I also wish to thank Dr. Francis Yeh for his careful reading and the comments on my thesis.

TABLE OF CONTENTS

Introduction	1
Part 1. Robust Jackknife Estimate	2
1.1. A Brief Review of Classical Regression and Robust Procedures	2
1.2. A Brief Review of Jackknife Procedure for Ordinary Regression	12
1.3. Jackknife Robust Estimate	18
Part 2. Simulation Study	38
2.1. The Model and The Data	38
2.2. Procedures	42
2.3. Measurement	49
2.4. The Simulation Results	50
2.5. Further Study	57
Part 3. Diagnostic Statistics for Robust Estimate	62
3.1. A brief Review of Diagnostics for Ordinary Least Squares Estimates	62
3.2. The Diagnostic Statistics of Robust Estimate	69
Part 4. Case Study	81
4.1. The Land Use and Water Quality Data	81
4.2. The Hertzsprung-Russell Star Data	89
References	93

LIST OF TABLES

1. The Bias of the Estimate 51

2. The Relative Bias of Variance-Covariance for Huber's M-estimate Case 53

3. The Relative Bias of Variance-Covariance for Mallows' GM-estimate Case 54

4. The Relative Bias of Variance-Covariance for LS and Schweppe's Estimates Case 55

5. The Recommended Procedures Based on the Relative Bias of Variance 56

6. Estimators of $\hat{q}$ for Huber's type M-estimate 59

7. Estimators of $\hat{q}$ for Mallows' type GM-estimate 60

8. Estimators of $\hat{q}$ for Schweppe's type GM-estimate 61

9. The Estimators of Land Data and Their Variance Estimation 83

10. The Estimators of Land Data and Their Variance Estimation (Case 5 Is Excluded) 88

LIST OF FIGURES

1. Scatterplot of Sample Data Generated by $\nu = 0.2$ 40

2. Scatterplot of Sample Data Generated by $\nu = 0$ 41

3. Normal Probability Plot for Residuals 84

4. Scatterplot of Hertzsproung-Russell Data 90

5. Plot of Diagnostic Statistics for Hertzsproung-Russell Data 91

Introduction

This thesis has four parts. In the first part, we discuss robust M and GM regression estimates. Since the jackknife procedure can be used to reduce the bias of an estimate as well as estimate the variance of the estimate, we combine the jackknife procedure and the robust procedure in our estimations. A delete-one jackknife procedure is applied in obtaining the robust estimates. Three types of jackknife estimates are considered, namely, the ordinary jackknife robust estimate, the weighted jackknife robust estimate and the general weighted robust estimate. The corresponding variance estimates are given as well. Furthermore, in consideration of the fact that the calculation of jackknife robust estimates is a time consuming procedure, a computationally efficient approximation procedure is proposed.

In the second part, we give some simulation results which are obtained based on the procedures we proposed in part one. A discussion on the further work on the inference of robust estimates is presented.

The third part deals with robust diagnostics. In this part, several diagnostic statistics are presented. Those diagnostic statistics are analogous to the diagnostic statistics in the ordinary least squares case but can be applied in robust procedures to detect influential observations and outliers.

In the fourth part, a case study based on the diagnostic statistics which are proposed in part three is conducted.

1 Robust Jackknife Estimate

1.1 A Brief Review of Classical Regression and Robust Procedures

1.1.1 The Classical Regression Procedure

Consider a regression problem with n cases $(y_i, \mathbf{x}_i^T)$. In matrix form, it can be expressed as

$$\mathbf{y} = X\beta + \varepsilon \tag{1.1}$$

$$X = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

where

$\mathbf{x}_i$ are non-random design points

β is a vector of parameters

ε is a vector of independent random variables with marginal distribution function F , and its expectation and variance are $E[\varepsilon] = \mathbf{0}$, and $Var(\varepsilon) = \sigma^2 I$

Classically, the problem of estimating β is solved by minimizing the sum of squares:

$$\sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \beta)^2 = min. \tag{1.2}$$

By taking partial derivatives, we see that the least squares etimator $\hat{\beta}_{LS}$ satisfies

$$\sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \hat{\beta}_{LS}) \mathbf{x}_i = \mathbf{0} \ , \tag{1.3}$$

so, the least squares estimator is

$$\hat{\beta}_{LS} = (X^T X)^{-1} X^T \mathbf{y}.$$

The variance of the least squares estimator is

$$Var(\hat{\beta}_{LS}) = \sigma^2 (X^T X)^{-1} .$$

The value of σ^2 can be estimated by the MSE (MSE stands for error mean square or residual mean square) which is defined by

$$MSE = \frac{1}{n-p} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \hat{\beta})^2 .$$

The estimate of the variance of the LS estimator is

$$v(\hat{\beta}_{LS}) = \widehat{Var(\hat{\beta}_{LS})} = MSE(X^T X)^{-1}$$

1.1.2 Robust Procedures for Regression Problems

1.1.2.1 Ordinary M-estimation

The classical equations (1.2) and (1.3) can be robustified in the following way. Instead of minimizing the sum of squares of residuals, we minimize a sum of less rapidly increasing functions of the residuals:

$$\sum_{i=1}^{n} \rho(y_i - \mathbf{x}_i^T \beta) = min. \tag{1.4}$$

After taking partial derivatives, we find that the robust M-estimator $\hat{\beta}_M$ should satisfy

$$\sum_{i=1}^{n} \psi(y_i - \mathbf{x}_i^T \hat{\beta}_M) \mathbf{x}_i = \mathbf{0} \tag{1.5}$$

where $\psi(\cdot) = \rho'(\cdot)$.

Assume that ρ is an even, convex function, so ψ is an odd, increasing function. To solve $\sum_{i=1}^{n} \psi(y_i - \mathbf{x}_i^T\beta)\mathbf{x}_i = 0$, we do weighted least squares regression iteratively to obtain the estimate. To solve this problem, we can write the above equation as

$$\sum_{i=1}^{n} \frac{\psi(y_i - \mathbf{x}_i^T\hat{\beta})}{y_i - \mathbf{x}_i^T\hat{\beta}}(y_i - \mathbf{x}_i^T\hat{\beta})\mathbf{x}_i = \mathbf{0} \ .$$

Put

$$w_i = \frac{\psi(y_i - \mathbf{x}_i^T\hat{\beta})}{y_i - \mathbf{x}_i^T\hat{\beta}} \ ,$$

and solve

$$\sum_{i=1}^{n} w_i(y_i - \mathbf{x}_i^T\hat{\beta})\mathbf{x}_i = \mathbf{0} \ .$$

In matrix form, this can be written as

$$X^T W\mathbf{y} = X^T W X\hat{\beta}$$

where $W = diag(w_1, \ldots, w_n)$.

For non-singular matrix $X^T WX$, we can obtain the estimator $\hat{\beta}$ as the solution to

$$\hat{\beta} = (X^T WX)^{-1} X^T W\mathbf{y} \ .$$

This is a weighted least squares problem but the weights depend on the parameters being estimated.

The algorithm is:

- Choose an initial estimate $\hat{\beta}_{(0)}$. Usually, we can choose the estimator $\hat{\beta}_{L1}$ which is obtained by using least absolute value method or the estimator $\hat{\beta}_{LS}$ which is obtained by ordinary least squares method.

- Calculate the weights

$$w_{i(0)} = \frac{\psi(y_i - \mathbf{x}_i^T \hat{\beta}_{(0)})}{y_i - \mathbf{x}_i^T \hat{\beta}_{(0)}}$$

and solve the equation

$$\sum_{i=1}^{n} w_{i(0)}(y_i - \mathbf{x}_i^T \hat{\beta}_{(1)})\mathbf{x}_i = \mathbf{0}$$

to obtain the weighted least squares estimator $\hat{\beta}_{(1)}$.

- Replace $\hat{\beta}_{(0)}$ by $\hat{\beta}_{(1)}$ and iterate to convergence.

Note: In most cases we need to consider the scale σ. This case will be mentioned later. We can use the same ideas to deal with such a problem.

For the simple form of M-estimate (*i.e.* the case in which we do not consider the scale), the asymptotic variance of the estimator is

$$Var(\hat{\beta}_M) = \frac{E(\psi(\varepsilon)^2)}{(E\psi'(\varepsilon))^2}(X^T X)^{-1} .$$

This can be estimated by

$$v(\hat{\beta}_M) = \frac{1/(n-p)\sum_{i=1}^{n} \psi^2(r_i)}{[(1/n)\sum_{i=1}^{n} \psi'(r_i)]^2}(X^T X)^{-1} ,$$

where $r_i = y_i - \mathbf{x}_i^T \hat{\beta}_M$.

Let the hat matrix $H = X(X^T X)^{-1} X^T$ have the diagonal $\{h_i\}_{i=1}^n$. For the balanced case (*i.e.* $h_i = h = p/n$), with symmetric error distributions and skew symmetric ψ function, assuming that $1 \ll p \ll n$, and if we neglect terms of the

orders $h^2 = (p/n)^2$ or $1/n$, Huber (1981) gives three expressions which are unbiased estimators of $Var(\hat{\beta}_M)$. The expressions are

$$\kappa^2 \frac{[1/(n-p)]\sum \psi^2(r_i)}{[(1/n)\sum \psi'(r_i)]^2}(X^TX)^{-1}$$

$$\kappa \frac{[1/(n-p)]\sum \psi^2(r_i)}{(1/n)\sum \psi'(r_i)}(X^TWX)^{-1}$$

$$\frac{1}{\kappa(n-p)}\sum \psi^2(r_i)(X^TWX)^{-1}(X^TX)(X^TWX)^{-1}$$

where

$$\kappa = 1 + \frac{p}{n}\frac{\text{var}(\psi')}{(E\psi')^2} ,$$

and

$$W = diag(\psi'(r_1), \ldots, \psi'(r_n)) .$$

The values of $E\psi'$ and $var(\psi')$ can be estimated by

$$E(\psi') \approx \frac{1}{n}\sum \psi'(r_i) =: m$$

$$var(\psi') \approx \frac{1}{n}\sum [\psi'(r_i) - m]^2 .$$

For the special case of Huber's ψ function

$$\psi(x) = \begin{cases} -k & x < -k \\ x & |x| < k \\ k & x > k \end{cases}$$

κ becomes

$$\kappa = 1 + \frac{p}{n}\frac{1-m}{m}$$

where m is the relative frequency of the residuals which satisfy $-k < r_i < k$.

1.1.2.2 Ordinary M-estimation of Regression and Scale

As mentioned before, for M-estimation we usually need to consider scale as well. In this case, formulas (1.4) and (1.5) become

$$\sum_{i=1}^{n} \rho(\frac{y_i - \mathbf{x}_i^T \beta}{\sigma}) = min. \tag{1.6}$$

and

$$\sum_{i=1}^{n} \psi(\frac{y_i - \mathbf{x}_i^T \beta}{\sigma})\mathbf{x}_i = \mathbf{0} \tag{1.7}$$

The scale $\sigma = \sigma(F)$ can be estimated by

$$\hat{\sigma} = \sigma(F_n) = MAD/\Phi^{-1}(3/4) \tag{1.8}$$

where MAD is the median absolute deviation (centred at 0) based on the full data set, and $\Phi(\cdot)$ represents the standard normal distribution function.

Note: If the errors are symmetrically distributed with standard deviation σ, then $\sigma(F) = \sigma F^{-1}(3/4)/\Phi^{-1}(3/4)$ ($= \sigma$ if $F = \Phi$) .

1.1.2.3 The Generalized M-estimate

It is well known that the solutions (the estimators) which satisfy formula (1.7) are only robust against outlying y values but not robust against highly influential $\mathbf{x}$ values. In order to improve our estimators, we can apply the Generalized M-estimate

which is robust against both types of deviations. The Generalized M-estimate $\widehat{\beta}_{GM}$ is obtained by solving

$$\sum_{i=1}^{n} \eta(\mathbf{x}_i, \frac{y_i - \mathbf{x}_i^T \beta}{\sigma})\mathbf{x}_i = \mathbf{0} \tag{1.9}$$

where the function $\eta(\mathbf{x}, \varepsilon)$ is an even function of $\mathbf{x}$ and an odd non-decreasing function of ε.

1.1.3 The General Form of Variance of Robust Estimate

For the M-estimate and Generalized M-estimate cases, we can summarize the expressions in a simple form $\phi(z_i, \beta)$. For the M-estimate case, we have

$$\phi(z_i, \beta) = \psi(\frac{y_i - \mathbf{x}_i^T \beta}{\sigma})\mathbf{x}_i \ ,$$

while for the GM-estimate case we have

$$\phi(z_i, \beta) = \eta(\mathbf{x}_i, \frac{y_i - \mathbf{x}_i^T \beta}{\sigma})\mathbf{x}_i \ ,$$

where $z_i = (\mathbf{x}_i^T, \ y_i)^T$.

To obtain the variance of the robust estimate, we can use the idea of the influence function which was introduced by Hampel (1968), (1974), Hampel et al (1986). The influence function of a statistic $T(F)$ at a point Z can be obtained by

$$IF(Z; T, F) = \frac{d}{d\varepsilon} T(F_\varepsilon)|_{\varepsilon=0, G=\Delta Z} \tag{1.10}$$

where

$$T(F_\varepsilon) = T((1-\varepsilon)F + \varepsilon G) \qquad (\ 0 \leq \varepsilon \leq 1\ ) \ ,$$

and the derivative of $T(F_\varepsilon)$ has the form:

$$\frac{d}{d\varepsilon}T(F_\varepsilon)|_{\varepsilon=0,G=\Delta Z} = \lim_{\varepsilon\to 0}\frac{T[(1-\varepsilon)F+\varepsilon\Delta Z]}{\varepsilon}\ .$$

Here $T(\cdot)$ is a vector-valued statistic based on a random sample from the cdf F and ΔZ is the distribution function of a random varible which equals Z with probability one.

Now we consider the influence function of the robust estimate. Let $\hat{\beta} = T(F_n)$ be the robust estimator based on sample size n which satisfies the general form of the robust esimate formula

$$\sum_{i=1}^{n}\phi(z_i, T(F_n)) = \mathbf{0} \tag{1.11}$$

where $z_1,\ldots,z_n$ are the values of the random variables $Z_1,\ldots,Z_n$ which have the distribution function F, and $F_n = \frac{1}{n}\sum_{i=1}^{n}\Delta z_i$ is the empirical distribution function. Put $T(F_\varepsilon) = T((1-\varepsilon)F+\varepsilon G)$, $0\le\varepsilon\le 1$. The influence function for robust estimator T on distribution function F at point z_i can be written as

$$IF(z_i;T,F) = \frac{d}{d\varepsilon}T(F_\varepsilon)|_{\varepsilon=0,G=\Delta z_i}\ .$$

After some calculation, we have

$$\sqrt{n}(T(F_n)-T(F)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}IF(z_i;T,F)+\sqrt{n}\cdot Rem\ .$$

If, as is generally the case, the remainder in the above equation is $o_p(n^{1/2})$, then by the Central Limit theorem, we have following expression,

$$\sqrt{n}(T(F_n)-T(F)) \xrightarrow{W} N(\mathbf{0}, E_F(IF\cdot IF^T))\ .$$

A calculation yields

$$IF(Z;T,F) = -(E_F[\phi'(Z,T(F))])^{-1}\phi(Z,T(F)) , \tag{1.12}$$

where

$$\phi'(Z,T(F)) = \frac{\partial}{\partial T}\phi(Z,T(F)) .$$

So the asymptotic variance of the estimator $\hat{\beta}$ can be obtained by

$$\begin{aligned}\lim_{n\to\infty} Var(\sqrt{n}\hat{\beta}) &= E_F(IF \cdot IF^T) \\ &= (E_F[\phi'(Z,T(F))])^{-1}E_F[\phi(Z,T(F)) \cdot \phi(Z,T(F))^T](E_F[\phi'(Z,T(F))])^{-1}.\end{aligned}$$

Write

$$M_\phi = -\sigma_\phi E_F[\phi'(Z,T(F))],$$

where $\sigma_\phi = \sigma(F)$ and write

$$Q_\phi = E_F[\phi(Z,T(F)) \cdot \phi(Z,T(F))^T] ,$$

Then the asymptotic variance of the estimate can be written as

$$\lim_{n\to\infty} Var(\sqrt{n}\hat{\beta}) = \sigma_\phi^2 M_\phi^{-1} Q_\phi M_\phi^{-1} \tag{1.13}$$

From the general form of function $\phi(\cdot,\cdot)$, we consider the variance estimate for robust M-estimate and GM-estimate respectively.

1.1.3.1 The Variance of M-estimate

In the case of M-estimate, the ϕ function is:

$$\phi(Z, T(F)) = \psi(\frac{y - \mathbf{x}^T\beta_M}{\sigma_M})\mathbf{x}$$

and

$$\phi'(Z, T(F)) = \frac{\partial}{\partial\beta_M}[\psi(\frac{y - \mathbf{x}^T\beta_M}{\sigma_M})\mathbf{x}] = -\psi'(\frac{y - \mathbf{x}^T\beta_M}{\sigma_M})\frac{\mathbf{x}\mathbf{x}^T}{\sigma_M}$$

so the variance of estimator can be obtained by

$$\lim_{n\to\infty} Var(\sqrt{n}\hat{\beta}_M) = \sigma_M^2 M_M^{-1} Q_M M_M^{-1} \tag{1.14}$$

where

$$M_M = \sigma_M E_F[\psi'(\frac{y - \mathbf{x}^T\beta_M}{\sigma_M})\frac{\mathbf{x}\mathbf{x}^T}{\sigma_M}] = E_F[\psi'(\frac{\varepsilon}{\sigma_M})\mathbf{x}\mathbf{x}^T]$$

and

$$Q_M = E_F[\psi^2(\frac{y - \mathbf{x}^T\beta_M}{\sigma_M})\mathbf{x}\mathbf{x}^T] = E_F[\psi^2(\frac{\varepsilon}{\sigma_M})\mathbf{x}\mathbf{x}^T] \ .$$

1.1.3.2 The Variance of GM-estimate

In the case of GM-estimate, the ϕ function is:

$$\phi(Z, T(F)) = \eta(\mathbf{x}, \frac{y - \mathbf{x}^T\beta_{GM}}{\sigma_{GM}})\mathbf{x}$$

and

$$\phi'(Z, T(F)) = \frac{\partial}{\partial\beta_{GM}}[\eta(\mathbf{x}, \frac{y - \mathbf{x}^T\beta_{GM}}{\sigma_{GM}})\mathbf{x}] = -\eta'(\mathbf{x}, \frac{y - \mathbf{x}^T\beta_{GM}}{\sigma_{GM}})\frac{\mathbf{x}\mathbf{x}^T}{\sigma_{GM}} \ .$$

The variance of GM-estimate can be written as

$$\lim_{n\to\infty} Var(\sqrt{n}\hat{\beta_{GM}}) = \sigma^2_{GM} M^{-1}_{GM} Q_{GM} M^{-1}_{GM} \tag{1.15}$$

where

$$M_{GM} = E_F[\eta'(\mathbf{x}, \frac{\varepsilon}{\sigma_{GM}})\mathbf{x}\mathbf{x}^T]$$

and

$$Q_{GM} = E_F[\eta^2(\mathbf{x}, \frac{\varepsilon}{\sigma_{GM}})\mathbf{x}\mathbf{x}^T] \ .$$

For above expressions of variance of M and GM-estimate, we can apply the method of moments to estimate M_M, Q_M, M_{GM}, and Q_{GM} and thus get estimates of $Var(\hat{\beta}_M)$ and $Var(\hat{\beta}_{GM})$.

1.2 A Brief Review of Jackknife Procedure for Ordinary Regression

1.2.1 About Jackknife Procedure

When we concern the nonparametric estimation of bias and the varaince, we can apply resample methods such as Jackknife, Bootstrap and Fisher's information theory (See Efron (1982)). In this paper, we consider applying Jackknife method in regression problem. Jackknife is a subsample replication technique. Quenouille (1949) originally introduced this procedure as a method of reducing the bias of an

estimator of a serial correlation coefficient. Later, in his paper (1956), Quenouille generalized the technique and explored its general bias reduction properties in an infinite population context. Research on the jackknife method divides basically along two distinct lines: 1). bias reduction and 2). variance estimation.

1.2.2 The Property of Jackknife

Suppose T_n is a biased estimator of the parameter θ in which we are interested. The jackknife estimator has the property (under certain conditions) that it removes the order $1/n$ term from the bias of the estimator (See Miller (1974a), Huber (1981)). Suppose the estimator T_n is computed from n *i.i.d.* observations so that T_{n-1} is computed in the same way from $n-1$ *i.i.d.* observations. The ith jackknifed pseudo-value is, by definition,

$$T_{ni}^{*} = nT_n - (n-1)T_{n-1}(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n).$$

If T_n is a consistent estimate of θ, whose bias has the asymptotic expression

$$E(T_n - \theta) = \frac{a_1}{n} + \frac{a_2}{n^2} + O(\frac{1}{n^3}) \,, \tag{1.16}$$

then $T_n^{*} = \frac{1}{n}\sum_i T_{ni}^{*}$ has a bias:

$$E(T_n^{*} - \theta) = -\frac{a_2}{n^2} + O(\frac{1}{n^3}) \,. \tag{1.17}$$

From the above expression, we know that the jackknife procedure can be used to reduce the bias of the estimator and improve the estimate.

1.2.3 The Application In Least Squares Regression

Let $\hat{\beta}$ be an estimator of the parameter β from the full sample. Let $\hat{\beta}_{(i)}$ be the estimator of the same functional form as $\hat{\beta}$, but computed from the reduced sample of size $n-1$ which is obtained by omitting the ith case.

1.2.3.1 Ordinary Jackknife In Classical Regression

When we neglect the unbalanced nature of the regression data, we can use the ordinary jackknife procedure to obtain the estimator for regression problem.

Define:

$$\hat{\beta}^i = n\hat{\beta} - (n-1)\hat{\beta}_{(i)} \qquad (i = 1, 2, \ldots, n) \; . \tag{1.18}$$

The $\hat{\beta}^i$ are called ordinary pseudovalues. The delete-one jackknife estimator is the mean of the pseudovalues $\hat{\beta}^i$

$$\hat{\beta}_J = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}^i \; . \tag{1.19}$$

For the ordinary jackknife estimator, by Tukey's suggestion (1958), its variance can be estimated by

$$v(\hat{\beta}_J) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\hat{\beta}^i - \hat{\beta}_J)(\hat{\beta}^i - \hat{\beta}_J)^T \tag{1.20}$$

The asymptotic properties of $\hat{\beta}_J$ and $v(\hat{\beta}_J)$ were studied by Miller (1974b) under strong conditions that excluded cases with very unbalanced design matrix X.

1.2.3.2 Weighted Jackknife In Classical Regression

If we consider the unbalanced nature of the regression data, we can use a modification of the pseudovalues.

Define

$$\hat{\beta}_w^i = \hat{\beta} + n(1-h_i)(\hat{\beta} - \hat{\beta}_{(i)}) \qquad (i = 1, 2, \ldots, n) \tag{1.21}$$

where $h_i = x_i^T(X^TX)^{-1}x_i$ is the diagonal elements of hat matrix. The $\hat{\beta}_w^i$ are called weighted pseudovalues, and were proposed by Hinkley (1977). The weighted jackknife estimator can be obtained by

$$\hat{\beta}_{J,w} = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_w^i = \hat{\beta} + \sum_{i=1}^{n}(1-h_i)(\hat{\beta} - \hat{\beta}_{(i)})$$

and the variance of weighted jackknife estimator can be obtained by

$$v(\hat{\beta}_{J,w}) = \frac{1}{n(n-p)}\sum_{i=1}^{n}(\hat{\beta}_w^i - \hat{\beta}_{J,w})(\hat{\beta}_w^i - \hat{\beta}_{J,w})^T \ .$$

For the least squares estimate, we know that

$$\hat{\beta} - \hat{\beta}_{(i)} = \frac{(X^TX)^{-1}\mathbf{x}_i r_i}{1-h_i} \tag{1.22}$$

The jackknife estimator becomes

$$\hat{\beta}_{J,w} = \hat{\beta} + (X^TX)^{-1}\sum_{i=1}^{n}\mathbf{x}_i r_i = \hat{\beta} \ . \tag{1.23}$$

Thus the weighted jackknife estimator $\hat{\beta}_{J,w}$ is identical to the original estimator $\hat{\beta}$ and the weighted jackknife variance becomes

$$v(\hat{\beta}_{J,w}) = \frac{1}{n(n-p)}\sum_{i=1}^{n}(\hat{\beta}_w^i - \hat{\beta})(\hat{\beta}_w^i - \hat{\beta})^T = \frac{n}{n-p}(X^TX)^{-1}(\sum_{i=1}^{n}r_i^2\mathbf{x}_i\mathbf{x}_i^T)(X^TX)^{-1} \ . \tag{1.24}$$

Where $v(\hat{\beta}_{J,w})$ is biased but is robust against error variance heterogeneity.

1.2.3.3 General Weighted Jackknife In Classical Regression

Wu (1986) proposed a more general jackknife method to improve the estimate. Let $\mathcal{S} = (\ i_1,\ldots,i_r\ )$ be a subset of $(1,2,\ldots,n)$, $\hat{\beta}_{\mathcal{S}}$ be the least square estimators of $(y_i, \mathbf{x}_i^T)$ for $i \in \mathcal{S}$. Then we have

$$\hat{\beta}_{\mathcal{S}} = (X_{\mathcal{S}}^T X_{\mathcal{S}})^{-1} X_{\mathcal{S}}^T \mathbf{y}_{\mathcal{S}} \ .$$

Let $\hat{\beta}$ be the least squares estimator for the full data set. Between $\hat{\beta}_{\mathcal{S}}$ and $\hat{\beta}$, we have

$$\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y} = \frac{\sum_r |X_{\mathcal{S}}^T X_{\mathcal{S}}| \hat{\beta}_{\mathcal{S}}}{\sum_r |X_{\mathcal{S}}^T X_{\mathcal{S}}|} \ .$$

where $\sum_r$ denote the summation over all the subsets $\mathcal{S}$ of size r.

The general weighted pseudovalues are proposed to be obtained by

$$\tilde{\beta}^{\mathcal{S}} = \hat{\beta} + (\frac{r-p+1}{n-r})^{1/2} (\hat{\beta}_{\mathcal{S}} - \hat{\beta}) \ .$$

The general weighted jackknife estimator can be obtained by

$$\hat{\beta}_{J,gw} = \sum_r \alpha_{\mathcal{S}} \tilde{\beta}^{\mathcal{S}} \ ,$$

where $\alpha_{\mathcal{S}}$ satisfy $\alpha_{\mathcal{S}} \propto |X_{\mathcal{S}}^T X_{\mathcal{S}}|$, and $\sum_r \alpha_{\mathcal{S}} = 1$.

The variance of above estimator is obtained by

$$v(\hat{\beta}_{J,gw}) = \sum_r \alpha_{\mathcal{S}} (\tilde{\beta}^{\mathcal{S}} - \hat{\beta})(\tilde{\beta}^{\mathcal{S}} - \hat{\beta})^T \ .$$

Substituting the general weighted pseudovalues $\tilde{\beta}_S$ in the above formula, the variance of the estimator becomes

$$v(\hat{\beta}_{J,gw}) = \frac{r-p+1}{n-r} \sum_r \alpha_S (\hat{\beta}_S - \hat{\beta})(\hat{\beta}_S - \hat{\beta})^T \ .$$

When $r = n-1$, this is the delete-one case. In this case, the general weighted pseudovalues are

$$\tilde{\beta}^i = \hat{\beta} + (n-p)^{1/2}(\hat{\beta}_{(i)} - \hat{\beta}) \qquad (i = 1, 2, \ldots, n) \tag{1.25}$$

The general weighted jackknife estimator is then

$$\hat{\beta}_{J,gw} = \sum_{i=1}^{n} \alpha_i \tilde{\beta}^i \ . \tag{1.26}$$

The corresponding variance estimate is

$$v(\hat{\beta}_{J,gw}) = (n-p) \sum_{i=1}^{n} \alpha_i (\hat{\beta}_{(i)} - \hat{\beta})(\hat{\beta}_{(i)} - \hat{\beta})^T \tag{1.27}$$

where $\alpha_i \propto |X_{(i)}^T X_{(i)}|$, and $\sum_r \alpha_i = 1$.

Since we have (see Wu (1986))

$$\sum_{i=1}^{n} |X_{(i)}^T X_{(i)}| = \begin{pmatrix} n-p \\ \\ 1 \end{pmatrix} |X^T X| = (n-p)|X^T X|$$

we can take

$$\alpha_i = (n-p)^{-1} \frac{|X_{(i)}^T X_{(i)}|}{|X^T X|} \ .$$

A standard result is that if a matrix A is invertible, then we have

$$|A - BC| = |A||I - CA^{-1}B| \ . \tag{1.28}$$

From this formula, we can obtain the following relation between $|X_{(i)}^T X_{(i)}|$ and $|X^T X|$:

$$|X_{(i)}^T X_{(i)}| = |X^T X - \mathbf{x}_i \mathbf{x}_i^T| = |X^T X|(1 - \mathbf{x}_i^T (X^T X)^{-1} \mathbf{x}_i) = |X^T X|(1 - h_i) \ .$$

So the weights can be written as

$$\alpha_i = (n-p)^{-1}(1-h_i)$$

The general weighted jackknife estimator is then

$$\hat{\beta}_{J,gw} = \sum_{i=1}^{n} (n-p)^{-1}(1-h_i)\tilde{\beta}^i = \beta - (n-p)^{-1/2}(X^T X)^{-1} \sum_{i=1}^{n} \mathbf{x}_i^T r_i = \hat{\beta} \qquad (1.29)$$

This shows that the general weighted jackknife estimator $\hat{\beta}_{J,gw}$ is also equal to the original estimator $\hat{\beta}$. The estimated variance of the general weighted jackknife estimator is

$$v(\hat{\beta}_{J,gw}) = \sum_{i=1}^{n} (1-h_i)(\hat{\beta}_{(i)} - \hat{\beta})(\hat{\beta}_{(i)} - \hat{\beta})^T = (X^T X)^{-1} (\sum_{i=1}^{n} \frac{r_i^2}{1-h_i} \mathbf{x}_i \mathbf{x}_i^T)(X^T X)^{-1} \ . \qquad (1.30)$$

Where $v(\hat{\beta}_{J,gw})$ is unbiased if the classical conditions are satisfied and is robust against error variance heteroscedasticity.

1.3 Jackknife Robust Estimate

The robust M and GM-estimates are biased in finite samples. In order to reduce the bias of the estimates and improve the estimates of the variances of the estimators, we propose to combine jackknife procedures with robust procedures. As we did in the previous section, we use the simplified expressions i.e. the general functional form

of $\phi(z_i, \hat{\beta}, \hat{\sigma})$ which can be used in both robust M and GM-estimate estimate cases. The problem of robust estimation becomes to solve the equations $\sum_i \phi(z_i, \beta, \sigma) = \mathbf{0}$. Let $\hat{\beta}_\phi$ be the estimator which satisfies

$$\sum_{j=1}^{n} \phi(z_j, \hat{\beta}_\phi, \hat{\sigma}_\phi) = \mathbf{0} \tag{1.31}$$

and $\hat{\beta}_{\phi(i)}$ $(i = 1, \ldots, n)$ be the estimators which satisfy

$$\sum_{j \neq i}^{n} \phi(z_j, \hat{\beta}_{\phi(i)}, \hat{\sigma}_{\phi(i)}) = \mathbf{0} \qquad (i = 1, \ldots, n)\ . \tag{1.32}$$

In other words, $\hat{\beta}_{\phi(i)}$ are the estimators of the same functional form as $\hat{\beta}_\phi$ but computed from the reduced sample of size $n-1$ obtained by omitting the ith case. For above formulas, $\hat{\sigma}_\phi$ can be obtained by (1.8), and $\hat{\sigma}_{\phi(i)}$ equal to $MAD_{(i)}/\Phi^{-1}(3/4)$. Since the MAD is insensitive to single deletions, we continue to use $\hat{\sigma}_\phi$, rather than $\hat{\sigma}_{\phi(i)}$. Formula (1.32) becomes

$$\sum_{j \neq i}^{n} \phi(z_j, \hat{\beta}_{\phi(i)}, \hat{\sigma}_\phi) = \mathbf{0} \qquad (i = 1, \ldots, n)\ . \tag{1.33}$$

1.3.1 The Exact Delete-one Jackknife Robust Estimate

For the above equations, we can use the algorithm mentioned before to obtain the estimators $\hat{\beta}_\phi$ and $\hat{\beta}_{\phi(i)}$ $(i = 1, \ldots, n)$. When we calculate these estimators, we need to go through the reweighted least squares method $(n+1)$ times. Each estimator will satisfy the corresponding equation exactly. We name those jackknife estimators which are obtained based on the $\hat{\beta}_\phi$ and $\hat{\beta}_{\phi(i)}$ $(i = 1, \ldots, n)$ the exact delete-one jackknife robust estimators.

1.3.1.1 The Ordinary Jackknife for Robust Estimation

From $\hat{\beta}_\phi$ and $\hat{\beta}_{\phi(i)}$ $(i = 1, \ldots, n)$, we define the ordinary pseudovalues as

$$\hat{\beta}_\phi^i = n\hat{\beta}_\phi - (n-1)\hat{\beta}_{\phi(i)} \ . \tag{1.34}$$

The delete-one jackknife estimator is the mean of $\hat{\beta}_\phi^i$:

$$\hat{\beta}_J^\phi = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_\phi^i \ . \tag{1.35}$$

Using Tukey's suggestion (1958), the variance of $\hat{\beta}$ can be estimated by

$$v(\hat{\beta}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\hat{\beta}_\phi^i - \hat{\beta}_J^\phi)(\hat{\beta}_\phi^i - \hat{\beta}_J^\phi)^T \tag{1.36}$$

Note: As mentioned by Tukey, it can also be used as an estimate of the variance of $\hat{\beta}_J^\phi$.

For the M-estimate, we have the ordinary pseudovalues

$$\hat{\beta}_M^i = n\hat{\beta}_M - (n-1)\hat{\beta}_{M(i)} \ . \tag{1.37}$$

In the above formula, the values $\hat{\beta}_{M(i)}$ satisfy

$$\sum_{j\neq i}^{n} \psi(r_{j(i)}^M/\hat{\sigma}_M)\mathbf{x}_j = 0 \ , \tag{1.38}$$

where

$$r_{j(i)}^M = y_j - \mathbf{x}_j^T\hat{\beta}_{M(i)} \ . \tag{1.39}$$

The corresponding variance is

$$v(\hat{\beta}_M) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\hat{\beta}_M^i - \hat{\beta}_J^M)(\hat{\beta}_M^i - \hat{\beta}_J^M)^T \ . \tag{1.40}$$

Similar to the case of M-estimate, we can obtained the ordinary pseudovalues for GM-estimate which are

$$\hat{\beta}_{GM}^i = n\hat{\beta}_{GM} - (n-1)\hat{\beta}_{GM(i)} \ . \tag{1.41}$$

The values $\hat{\beta}_{GM(i)}$ in above formula are satisfy

$$\sum_{j \neq i}^{n} \eta(\mathbf{x}_j, r_{j(i)}^{GM} / \hat{\sigma}_{GM}) \mathbf{x}_j = 0 \ , \tag{1.42}$$

where

$$r_{j(i)}^{GM} = y_j - \mathbf{x}_j^T \hat{\beta}_{M(i)} \ .$$

The estimation of the variance is

$$v(\hat{\beta}_{GM}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\hat{\beta}_{GM}^i - \hat{\beta}_J^{GM})(\hat{\beta}_{GM}^i - \hat{\beta}_J^{GM})^T \ . \tag{1.43}$$

1.3.1.2 The Weighted Delete-one Jackknife Robust Estimate

Similar to the weights which Hinkley proposed (1977) for ordinary least squares regression, we propose to use the weights

$$w_i^{\phi} = tr\{[-\frac{\partial}{\partial \hat{\beta}_{\phi}} \phi(z_i, \hat{\beta}_{\phi}, \hat{\sigma}_{\phi})] \hat{\sigma}_{\phi} \widehat{M}_{\phi}^{-1}\} \tag{1.44}$$

where

$$\widehat{M}_{\phi} = -\hat{\sigma}_{\phi} \sum_{i=1}^{n} \frac{\partial}{\partial \hat{\beta}_{\phi}} \phi(z_i, \hat{\beta}_{\phi}, \hat{\sigma}_{\phi}) \ .$$

We call the w_i^{ϕ} *robust leverages* for the robust estimates.

We can see that the w_i^{ϕ} are analogous to the weights for the ordinary least squares case, and are identical to these weights when least squares is used. In the ordinary least squares case, we have

$$\sum_{i=1}^{n} h_i = p \ .$$

For w_i^{ϕ}, we also have

$$\begin{aligned} \sum_{i=1}^{n} w_i^{\phi} &= \sum_{i=1}^{n} tr\{[-\frac{\partial}{\partial \hat{\beta}_{\phi}} \phi(z_i, \hat{\beta}_{\phi}, \hat{\sigma}_{\phi})] \hat{\sigma}_{\phi} \widehat{M}_{\phi}^{-1}\} \\ &= tr\{[-\hat{\sigma}_{\phi} \sum_{i=1}^{n} \frac{\partial}{\partial \hat{\beta}_{\phi}} \phi(z_i, \hat{\beta}_{\phi}, \hat{\sigma}_{\phi})] \widehat{M}_{\phi}^{-1}\} \\ &= tr[I_p] \ = \ p \ . \end{aligned}$$

For the OLS case, we have

$$\phi(z_i, \hat{\beta}_{\phi}, \hat{\sigma}_{\phi}) = \mathbf{x}_i (y_i - \mathbf{x}_i^T \hat{\beta}_{\phi}) \ ,$$

so

$$\frac{\partial}{\partial \hat{\beta}_{\phi}} \phi(z_i, \hat{\beta}_{\phi}, \hat{\sigma}_{\phi}) = -\mathbf{x}_i \mathbf{x}_i^T \ ,$$

$\widehat{M}_{\phi}$ can be expressed as

$$\widehat{M}_{\phi} = \hat{\sigma}_{\phi} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T = \hat{\sigma}_{\phi} X^T X \ ,$$

and w_i^{ϕ} becomes

$$w_i^{\phi} = tr\{[\mathbf{x}_i \mathbf{x}_i^T](X^T X)^{-1}\} = h_i \ .$$

The weighted pseudovalues for robust estimation are defined by

$$\hat{\beta}_{\phi,w}^{i} = \hat{\beta}_{\phi} + n(1 - w_i^{\phi})(\hat{\beta}_{\phi} - \hat{\beta}_{\phi(i)}) \ . \tag{1.45}$$

The weighted delete-one jackknife robust estimate is then

$$\hat{\beta}^{\phi}_{J,w} = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}^{i}_{\phi,w} \ . \tag{1.46}$$

The estimated variance of the estimator is

$$v(\hat{\beta}^{\phi}_{J,w}) = \frac{1}{n(n-p)}\sum_{i=1}^{n}(\hat{\beta}^{i}_{\phi,w} - \hat{\beta}^{\phi}_{J,w})(\hat{\beta}^{i}_{\phi,w} - \hat{\beta}^{\phi}_{J,w})^{T} \ . \tag{1.47}$$

In the case of M-estimation, the weights (robust leverages) become

$$w_i^M = tr\{-\frac{\partial}{\partial\hat{\beta}_M}[\psi(\frac{y_i - \mathbf{x}_i^T\hat{\beta}_M}{\hat{\sigma}_M})\mathbf{x}_i]\hat{\sigma}_M\widehat{M}_M^{-1}\} \tag{1.48}$$

$$= \psi'(\frac{r_i^M}{\hat{\sigma}_M})\mathbf{x}_i^T\widehat{M}_M^{-1}\mathbf{x}_i \tag{1.49}$$

where

$$\widehat{M}_M = \sum_{i=1}^{n}\psi'(\frac{r_i^M}{\hat{\sigma}_M})\mathbf{x}_i\mathbf{x}_i^T \ , \tag{1.50}$$

and

$$r_i^M = y_i - \mathbf{x}_i^T\hat{\beta}_M \ .$$

Note: If we let

$$\mathbf{u}_i = \sqrt{\psi'(\frac{r_i^M}{\hat{\sigma}_M})}\ \mathbf{x}_i \qquad\qquad (\ i,\ldots,n\ )$$

then we have

$$\widehat{M}_M = \sum_{i=1}^{n}\mathbf{u}_i\mathbf{u}_i^T = U^TU \ .$$

Define

$$W^M = U(U^TU)^{-1}U^T \ .$$

Then $diag(W^M) = (w_1^M,\ldots,w_n^M)$ and W^M is an idempotent matrix which plays the same role in robust estimation as the hat matrix H in ordinary regression.

Now we consider the weighted jackknife M-estimate. From (1.45), the weighted pseudovalues are

$$\hat{\beta}^i_{M,w} = \hat{\beta}_M + n(1 - w_i^M)(\hat{\beta}_M - \hat{\beta}_{M(i)}) \ . \tag{1.51}$$

The weighted delete-one jackknife M-estimator is

$$\hat{\beta}^M_{J,w} = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}^i_{M,w} \ . \tag{1.52}$$

The estimate of the variance of the jackknife estimator is

$$v(\hat{\beta}^M_{J,w}) = \frac{1}{n(n-p)}\sum_{i=1}^{n}(\hat{\beta}^i_{M,w} - \hat{\beta}^M_{J,w})(\hat{\beta}^i_{M,w} - \hat{\beta}^M_{J,w})^T \ . \tag{1.53}$$

Similar to the M-estimate, in the case of GM-estimate, the weights are

$$w_i^{GM} = \eta'(\mathbf{x}_i, \frac{r_i^{GM}}{\hat{\sigma}_{GM}})\mathbf{x}_i^T \widehat{M}_{GM}^{-1}\mathbf{x}_i \ , \tag{1.54}$$

where

$$\widehat{M}_{GM} = \sum_{i=1}^{n} \eta'(\mathbf{x}_i, \frac{r_i^{GM}}{\hat{\sigma}_{GM}})\mathbf{x}_i\mathbf{x}_i^T \ , \tag{1.55}$$

and

$$r_i^{GM} = y_i - \mathbf{x}_i^T\hat{\beta}_{GM} \ .$$

The weighted pseudovalues are

$$\hat{\beta}^i_{GM,w} = \hat{\beta}_{GM} + n(1 - w_i^{GM})(\hat{\beta}_{GM} - \hat{\beta}_{GM(i)}) \ . \tag{1.56}$$

The weighted delete-one jackknife GM-estimator can be obtained by

$$\hat{\beta}^{GM}_{J,w} = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}^i_{GM,w} \ . \tag{1.57}$$

The variance of the delete-one GM-estimate is

$$v(\hat{\beta}^{GM}_{J,w}) = \frac{1}{n(n-p)} \sum_{i=1}^{n} (\hat{\beta}^{i}_{GM,w} - \hat{\beta}^{GM}_{J,w})(\hat{\beta}^{i}_{GM,w} - \hat{\beta}^{GM}_{J,w})^T \ . \tag{1.58}$$

1.3.1.3 The General Weighted Delete-one Jackknife Robust Estimate

For the general weighted delete-one jackknife robust estimate, we use the pseudovalues as

$$\hat{\beta}^{i}_{\phi,gw} = \hat{\beta}_{\phi} + (n-p)^{1/2}(\hat{\beta}_{\phi(i)} - \hat{\beta}_{\phi}). \tag{1.59}$$

The general weighted jackknife estimator can be obtained by

$$\hat{\beta}^{\phi}_{J,gw} = \sum_{i=1}^{n} \alpha^{\phi}_{i} \hat{\beta}^{i}_{\phi,gw} \ , \tag{1.60}$$

where

$$\alpha^{\phi}_{i} = (n-p)^{-1}(1 - w^{\phi}_{i}) \ .$$

The variance of the jackknife estimator is

$$v(\hat{\beta}^{\phi}_{J,gw}) = \sum_{i=1}^{n} \alpha^{\phi}_{i} (\hat{\beta}^{i}_{\phi,gw} - \hat{\beta}^{\phi}_{J,gw})(\hat{\beta}^{i}_{\phi,gw} - \hat{\beta}^{\phi}_{J,gw})^T \ . \tag{1.61}$$

From above formulas, we can get the general weighted jackknife estimators and the corresponding variance for the cases of M and GM-estimate. For the M-estimate, the pseudovalues are

$$\hat{\beta}^{i}_{M,gw} = \hat{\beta}_{M} + (n-p)^{1/2}(\hat{\beta}_{M(i)} - \hat{\beta}_{M}) \tag{1.62}$$

The general weighted jackknife estimator for M-estimate is

$$\hat{\beta}^{M}_{J,gw} = \sum_{i=1}^{n} \alpha_i^M \hat{\beta}^{i}_{M,gw} \ , \tag{1.63}$$

where

$$\alpha_i^M = (n-p)^{-1}(1 - w_i^M) \ .$$

The variance of the jackknife estimator is

$$v(\hat{\beta}^{M}_{J,gw}) = \sum_{i=1}^{n} \alpha_i^M (\hat{\beta}^{i}_{M,gw} - \hat{\beta}^{M}_{J,gw})(\hat{\beta}^{i}_{M,gw} - \hat{\beta}^{M}_{J,gw})^T \ . \tag{1.64}$$

For GM-estimate case, the pseudovalues are

$$\hat{\beta}^{i}_{GM,gw} = \hat{\beta}_{GM} + (n-p)^{1/2}(\hat{\beta}_{GM(i)} - \hat{\beta}_{GM}) \ . \tag{1.65}$$

The general weighted jackknife estimator for GM-estimate is

$$\hat{\beta}^{GM}_{J,gw} = \sum_{i=1}^{n} \alpha_i^{GM} \hat{\beta}^{i}_{GM,gw} \ , \tag{1.66}$$

where

$$\alpha_i^{GM} = (n-p)^{-1}(1 - w_i^{GM}) \ .$$

The variance of the estimator is

$$v(\hat{\beta}^{GM}_{J,gw}) = \sum_{i=1}^{n} \alpha_i^{GM} (\hat{\beta}^{i}_{GM,gw} - \hat{\beta}^{GM}_{J,gw})(\hat{\beta}^{i}_{GM,gw} - \hat{\beta}^{GM}_{J,gw})^T \tag{1.67}$$

1.3.2 The Approximate Delete-one Jackknife Robust Estimate

1.3.2.1 Motivation

A common way to calculate the robust M or GM-estimate is to do weighted least squares regression iteratively. This is a time consuming procedure if we calculate the exact jackknife robust estimators, especially when the number of cases is large. In order to reduce the computing time, we can use an approximate process to calculate the delete-one jackknife robust estimators. Under this procedure, we only need to calculate one estimator which is obtained based on the full data set. From this estimator, we can get the the other estimators which are considered as the estimators of delete ith case $(i = 1, \ldots, n)$. By using this approximate procedure, the computing time is substantially reduced.

1.3.2.2 Approach

Suppose $\hat{\beta}_\phi$ and $\hat{\beta}_{\phi(i)}$ $(i = 1, \ldots, n)$ are the robust estimators which satisfy formulas (1.31) and (1.33) respectively. Let $\delta_{\phi(i)} = \hat{\beta}_\phi - \hat{\beta}_{\phi(i)}$. Then we can write

$$\hat{\beta}_{\phi(i)} = \hat{\beta}_\phi - \delta_{\phi(i)} \,. \tag{1.68}$$

Put (1.68) in (1.33), then by Taylor's expansion we have

$$0 = \sum_{j \neq i}^{n} \phi(z_j, \hat{\beta}_\phi - \delta_{\phi(i)}, \hat{\sigma}_\phi) \approx \sum_{j \neq i}^{n} \phi(z_j, \hat{\beta}_\phi, \hat{\sigma}_\phi) - \sum_{j \neq i}^{n} \frac{\partial}{\partial \hat{\beta}_\phi} [\phi(z_j, \hat{\beta}_\phi, \hat{\sigma}_\phi)] \delta_{\phi(i)} \,. \tag{1.69}$$

From (1.31), we have

$$\sum_{j\neq i}^{n} \phi(z_j, \hat{\beta}_\phi, \hat{\sigma}_\phi) = -\phi(z_i, \hat{\beta}_\phi, \hat{\sigma}_\phi) \ . \tag{1.70}$$

Combining (1.69) and (1.70), we obtain

$$\delta_{\phi(i)} \approx -(\sum_{j\neq i}^{n} \phi'(z_j, \hat{\beta}_\phi, \hat{\sigma}_\phi))^{-1} \phi(z_i, \hat{\beta}_\phi, \hat{\sigma}_\phi) \ , \tag{1.71}$$

where

$$\phi'(z_j, \hat{\beta}_\phi, \hat{\sigma}_\phi) = \frac{\partial}{\partial \hat{\beta}_\phi} \phi(z_j, \hat{\beta}_\phi, \hat{\sigma}_\phi) \ .$$

From above formulas, we can get the general form of approximate robust estimators $\hat{\beta}_{\phi(i)}$, $(1, \ldots, n)$. The expressions are

$$\begin{aligned} \hat{\beta}_{\phi(i)} &\approx \hat{\beta}_\phi + (\sum_{j\neq i}^{n} \phi'(z_j, \hat{\beta}_\phi, \hat{\sigma}_\phi))^{-1} \phi(z_i, \hat{\beta}_\phi, \hat{\sigma}_\phi) \\ &= \hat{\beta}_\phi - (\frac{\widehat{M}_\phi}{\hat{\sigma}_\phi} + \phi'(z_i, \hat{\beta}_\phi, \hat{\sigma}_\phi))^{-1} \phi(z_i, \hat{\beta}_\phi, \hat{\sigma}_\phi) \ . \end{aligned}$$

We denote $\hat{\beta}^a_{\phi(i)}$ as the approximation of $\hat{\beta}_{\phi(i)}$ which has the expression

$$\hat{\beta}^a_{\phi(i)} = \hat{\beta}_\phi - (\frac{\widehat{M}_\phi}{\hat{\sigma}_\phi} + \phi'(z_i, \hat{\beta}_\phi, \hat{\sigma}_\phi))^{-1} \phi(z_i, \hat{\beta}_\phi, \hat{\sigma}_\phi) \ . \tag{1.72}$$

It will be seen that this approximation makes the process of calculation much easier in solving the robust jackknife estimate.

Note: If we consider the one step Newton method to solve for $\hat{\beta}_{\phi(i)}$ in the equation

$$\sum_{j\neq i}^{n} \phi(z_j, \hat{\beta}_{\phi(i)}, \hat{\sigma}_\phi) = \mathbf{0}$$

starting with $\hat{\beta}_\phi$, we have that

$$\mathbf{0} = \sum_{j\neq i}^{n} \phi(z_j, \hat{\beta}_{\phi(i)}, \hat{\sigma}_\phi) \approx \sum_{j\neq i}^{n} \phi(z_j, \hat{\beta}_\phi, \hat{\sigma}_\phi) + [\frac{\partial}{\partial \hat{\beta}_\phi}\{\sum_{j\neq i}^{n} \phi(z_j, \hat{\beta}_\phi, \hat{\sigma}_\phi)\}](\hat{\beta}_{\phi(i)} - \hat{\beta}_\phi) \ .$$

Thus

$$\hat{\beta}_{\phi(i)} \approx \hat{\beta}_\phi + [\frac{\partial}{\partial \hat{\beta}_\phi}\{\sum_{j\neq i}^{n} \phi(z_j, \hat{\beta}_\phi, \hat{\sigma}_\phi)\}]^{-1} \phi(z_i, \hat{\beta}_\phi, \hat{\sigma}_\phi) = \hat{\beta}^a_{\phi(i)} \ .$$

Comparing this result with formula (1.72), we see that the approximate procedure we used here is essentially obtained by doing one step of Newton's method, starting with $\hat{\beta}_\phi$.

From formula (1.72), we can get the delete-one estimators of M-estimate and GM-estimate cases for using approximate procedure. For the M-estimate case, the approximate estimators for delete ith case are

$$\hat{\beta}^a_{M(i)} = \hat{\beta}_M - (\frac{\widehat{M}_M}{\hat{\sigma}_M} - \frac{1}{\hat{\sigma}_M}\psi'(\frac{r_i^M}{\hat{\sigma}_M})\mathbf{x}_i\mathbf{x}_i^T)^{-1}\psi(\frac{r_i^M}{\hat{\sigma}_M})\mathbf{x}_i \ .$$

After some algebra, we can obtain

$$\hat{\beta}^a_{M(i)} = \hat{\beta}_M - \hat{\sigma}_M \frac{\psi(r_i^M/\hat{\sigma}_M)}{1 - w_i^M} \widehat{M}_M^{-1}\mathbf{x}_i \ . \tag{1.73}$$

Similar to M-estimate case, we can obtain the formula for approximate delete-one GM-estimate. The approximate estimators of delete ith $\hat{\beta}^a_{GM(i)}$ can be gotten by

$$\hat{\beta}^a_{GM(i)} = \hat{\beta}_{GM} - \hat{\sigma}_{GM} \frac{\eta(\mathbf{x}_i, r_i^{GM}/\hat{\sigma}_{GM})}{1 - w_i^{GM}} \widehat{M}_{GM}^{-1}\mathbf{x}_i \ . \tag{1.74}$$

From formulas (1.73) and (1.74), we see that we need only obtain one robust estimate, then we can simply do some matrix calculation to get the rest of the approximate delete-one robust estimators. Using the above procedure, we can save a lot of computing time.

1.3.2.3 The Approximate Jackknife Estimate

Similar to the jackknife estimate which is mentioned in the previous section (the exact delete-one jackknife robust estimate), we can consider the jackknife estimate based on estimators $\hat{\beta}_\phi$ and $\hat{\beta}^a_{\phi(i)}$ $(i = 1, \ldots, n)$ which are obtained by using the approximate procedure. We name the jackknife estimate which is obtained by $\hat{\beta}_\phi$ and $\hat{\beta}^a_{\phi(i)}$ $(i = 1, \ldots, n)$ as approximate delete-one jackknife robust estimate. For the ordinary, weighted and general weighted jackknife estimate, we can use following formulas.

(1). Ordinary Approximate Jackknife Estimate

For the ordinary approximate jackknife estimate, the ordinary pseudovalues are

$$\hat{\beta}^{i,a}_\phi = n\hat{\beta}_\phi - (n-1)\hat{\beta}^a_{\phi(i)} = \hat{\beta}_\phi + (n-1)(\hat{\beta}_\phi - \hat{\beta}^a_{\phi(i)}) \,. \tag{1.75}$$

The approximate jackknife robust estimator is

$$\hat{\beta}^{\phi,a}_J = \hat{\beta}_\phi + \frac{n-1}{n}\sum_{i=1}^{n}(\hat{\beta}_\phi - \hat{\beta}^a_{\phi(i)}) \,. \tag{1.76}$$

The estimate of variance of jackknife estimator is

$$v(\hat{\beta}_J^{\phi,a}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\hat{\beta}_\phi^{i,a} - \hat{\beta}_J^{\phi,a})(\hat{\beta}_\phi^{i,a} - \hat{\beta}_J^{\phi,a})^T \ . \tag{1.77}$$

From above formulas, we can get the ordinary jackknife estimators for M and GM-estimation cases when we use corresponding ψ or η function to replace the ϕ function.

For the M-estimate, the ordinary pseudovalues of the approximate jackknife estimator are

$$\hat{\beta}_M^{i,a} = \hat{\beta}_M + (n-1)(\hat{\beta}_M - \hat{\beta}_{M(i)}^a) \ . \tag{1.78}$$

The approximate jackknife M-estimator is

$$\hat{\beta}_J^{M,a} = \hat{\beta}_M + \frac{n-1}{n} \sum_{i=1}^{n} (\hat{\beta}_M - \hat{\beta}_{M(i)}^a) \ . \tag{1.79}$$

The corresponding estimate of variance is

$$v(\hat{\beta}_J^{M,a}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\hat{\beta}_M^{i,a} - \hat{\beta}_J^{M,a})(\hat{\beta}_M^{i,a} - \hat{\beta}_J^{M,a})^T \ . \tag{1.80}$$

For the GM-estimate, the ordinary pseudovalues are

$$\hat{\beta}_{GM}^{i,a} = \hat{\beta}_{GM} + (n-1)(\hat{\beta}_{GM} - \hat{\beta}_{GM(i)}^a) \ . \tag{1.81}$$

the approximate jackknife GM-estimator is

$$\hat{\beta}_J^{GM,a} = \hat{\beta}_{GM} + \frac{n-1}{n} \sum_{i=1}^{n} (\hat{\beta}_{GM} - \hat{\beta}_{GM(i)}^a) \ . \tag{1.82}$$

The corresponding estimate of variance is

$$v(\hat{\beta}_J^{GM,a}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\hat{\beta}_{GM}^{i,a} - \hat{\beta}_J^{GM,a})(\hat{\beta}_{GM}^{i,a} - \hat{\beta}_J^{GM,a})^T \ . \tag{1.83}$$

(2). Weighted Approximate Jackknife Estimate

In this case, the weighted pseudovalues of approximate procedure $\hat{\beta}_{\phi}^{i,a}$ can be obtained by

$$\hat{\beta}_{\phi}^{i,a} = \hat{\beta}_{\phi} + n(1 - w_i^{\phi})(\hat{\beta}_{\phi} - \hat{\beta}_{\phi(i)}^{a}) \ . \tag{1.84}$$

The weighted jackknife robust estimator is

$$\hat{\beta}_{J,w}^{\phi,a} = \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}_{\phi}^{i,a} = \hat{\beta}_{\phi} + \sum_{i=1}^{n} (1 - w_i^{\phi})(\hat{\beta}_{\phi} - \hat{\beta}_{\phi(i)}^{a}) \ . \tag{1.85}$$

The variance of the estimator is

$$v(\hat{\beta}_{J,w}^{\phi,a}) = \frac{1}{n(n-p)} \sum_{i=1}^{n} (\hat{\beta}_{\phi}^{i,a} - \hat{\beta}_{J,w}^{\phi,a})(\hat{\beta}_{\phi}^{i,a} - \hat{\beta}_{J,w}^{\phi,a})^T \ . \tag{1.86}$$

For the M-estimate, we have the weighted pseudovalues

$$\hat{\beta}_M^{i,a} = \hat{\beta}_M + n(1 - w_i^M)(\hat{\beta}_M - \hat{\beta}_{M(i)}^{a}) \ . \tag{1.87}$$

The weighted jackknife M-estimator of approximate procedure can be written as

$$\hat{\beta}_{J,w}^{M,a} = \hat{\beta}_M + \sum_{i=1}^{n} (1 - w_i^M)(\hat{\beta}_M - \hat{\beta}_{M(i)}^{a}) \ . \tag{1.88}$$

Since

$$\sum_{i=1}^{n} (1 - w_i^M)(\hat{\beta}_M - \hat{\beta}_{M(i)}^{a}) = \sum_{i=1}^{n} (1 - w_i^M) \hat{\sigma}_M \frac{\widehat{M}_M^{-1}}{1 - w_i^M} \psi(\frac{r_i^M}{\hat{\sigma}_M}) \mathbf{x}_i = \hat{\sigma}_M \widehat{M}_M^{-1} \sum_{i=1}^{n} \psi(\frac{r_i^M}{\hat{\sigma}_M}) \mathbf{x}_i = \mathbf{0},$$

we have that the weighted approximate jackknife estimator $\hat{\beta}_{J,w}^{M,a}$ is identical to the original robust estimator $\hat{\beta}_M$. The variance of the estimator can be written as

$$\begin{aligned} v(\hat{\beta}_{J,w}^{M,a}) &= \frac{1}{n(n-p)} \sum_{i=1}^{n} (\hat{\beta}_M^{i,a} - \hat{\beta}_{J,w}^{M,a})(\hat{\beta}_M^{i,a} - \hat{\beta}_{J,w}^{M,a})^T \\ &= \frac{n}{n-p} \sum_{i=1}^{n} (1 - w_i^M)^2 (\hat{\beta}_M - \hat{\beta}_{M(i)}^a)(\hat{\beta}_M - \hat{\beta}_{M(i)}^a)^T \\ &= \frac{n}{n-p} \hat{\sigma}_M^2 \sum_{i=1}^{n} \psi^2(\frac{r_i^M}{\hat{\sigma}_M}) \widehat{M}_M^{-1} \mathbf{x}_i \mathbf{x}_i^T \widehat{M}_M^{-1} \ . \end{aligned}$$

We can write the above formula in the form

$$v(\hat{\beta}_{J,w}^{M,a}) = \frac{n}{n-p} \hat{\sigma}_M^2 \widehat{M}_M^{-1} \widehat{Q}_M \widehat{M}_M^{-1} \ , \tag{1.89}$$

where

$$\widehat{Q}_M = \sum_{i=1}^{n} \psi^2(\frac{r_i^M}{\hat{\sigma}_M}) \mathbf{x}_i \mathbf{x}_i^T \ . \tag{1.90}$$

For the case of GM-estimate, we have the weighted pseudovalues

$$\hat{\beta}_{GM}^{i,a} = \hat{\beta}_{GM} + n(1 - w_i^{GM})(\hat{\beta}_{GM} - \hat{\beta}_{GM(i)}^a) \ . \tag{1.91}$$

Similarly, we can get that the weighted approximate jackknife estimator $\hat{\beta}_{J.w}^{GM.a}$ is identical to the original robust estimator $\hat{\beta}_{GM}$ also. The corresponding variance of the jackknife estimator is

$$v(\hat{\beta}_{J,w}^{GM,a}) = \frac{n}{n-p} \hat{\sigma}_{GM}^2 \widehat{M}_{GM}^{-1} \widehat{Q}_{GM} \widehat{M}_{GM}^{-1} \ , \tag{1.92}$$

where

$$\widehat{Q}_{GM} = \sum_{i=1}^{n} \eta^2(\mathbf{x}_i, \frac{r_i^{GM}}{\hat{\sigma}_{GM}}) \mathbf{x}_i \mathbf{x}_i^T \ . \tag{1.93}$$

(3). General Weighted Approximate Jackknife Estimate

Applying the general weighted jackknife method to the approximate procedure, we define the pseudovalues as

$$\hat{\beta}^{i,a}_{\phi,gw} = \hat{\beta}_{\phi} + (n-p)^{1/2}(\hat{\beta}^{a}_{\phi(i)} - \hat{\beta}_{\phi}) \ . \tag{1.94}$$

The general weighted jackknife estimator can be obtained by

$$\hat{\beta}^{\phi,a}_{J,gw} = \sum_{i=1}^{n} \alpha_i^{\phi} \hat{\beta}^{i,a}_{\phi,gw} \ . \tag{1.95}$$

The variance of the general weighted jackknife estimator is

$$v(\hat{\beta}^{\phi,a}_{J,gw}) = \sum_{i=1}^{n} \alpha_i^{\phi} (\hat{\beta}^{i,a}_{\phi,gw} - \hat{\beta}^{\phi,a}_{J,gw})(\hat{\beta}^{i,a}_{\phi,gw} - \hat{\beta}^{\phi,a}_{J,gw})^T \ , \tag{1.96}$$

where

$$\alpha_i^{\phi} = (n-p)^{-1}(1 - w_i^{\phi}) \ .$$

In the case of M-estimate, the general weighted pseudovalues are

$$\hat{\beta}^{i,a}_{M,gw} = \hat{\beta}_{M} + (n-p)^{1/2}(\hat{\beta}^{a}_{M(i)} - \hat{\beta}_{M}) \ . \tag{1.97}$$

The general weighted jackknife for approximate procedure is

$$\hat{\beta}^{M,a}_{J,gw} = \sum_{i=1}^{n} \alpha_i^{M} \hat{\beta}^{i,a}_{M,gw} = \hat{\beta}_{M} + (n-p)^{-1/2} \sum_{i=1}^{n} (1 - w_i^{M})(\hat{\beta}^{a}_{M(i)} - \hat{\beta}_{M}) = \hat{\beta}_{M} \ .$$

That means the general weighted jackknife estimator $\hat{\beta}^{M,a}_{J,gw}$ is identical to the original robust estimator $\hat{\beta}_{M}$. The variance of the estimator is

$$\begin{aligned} v(\hat{\beta}^{M,a}_{J,gw}) &= \sum_{i=1}^{n} \alpha_i^{M} (\hat{\beta}^{i,a}_{M,gw} - \hat{\beta}_{M})(\hat{\beta}^{i,a}_{M,gw} - \hat{\beta}_{M})^T \\ &= \sum_{i=1}^{n} (1 - w_i^{M})(\hat{\beta}^{a}_{M(i)} - \hat{\beta}_{M})(\hat{\beta}^{a}_{M(i)} - \hat{\beta}_{M})^T \ , \end{aligned}$$

which can be simplified as

$$v(\hat{\beta}_{J,gw}^{M,a}) = \hat{\sigma}_M^2 \widehat{M}_M^{-1} \hat{Q}_{M,gw} \widehat{M}_M^{-1} \ , \tag{1.98}$$

where

$$\hat{Q}_{M,gw} = \sum_{i=1}^{n} \frac{\psi^2(r_i^M/\hat{\sigma}_M)}{1 - w_i^M} \mathbf{x}_i \mathbf{x}_i^T \ . \tag{1.99}$$

In the GM-estimate case, the general weighted pseudovalues are

$$\hat{\beta}_{GM,gw}^{i,a} = \hat{\beta}_{GM} + (n-p)^{1/2}(\hat{\beta}_{GM(i)}^{a} - \hat{\beta}_{GM}) \ . \tag{1.100}$$

In the similar manner, we can see that the estimator $\hat{\beta}_{J,gw}^{GM,a}$ is also identical to the original robust estimator $\hat{\beta}_{GM}$. The variance of the estimator can be written as

$$v(\hat{\beta}_{J,gw}^{GM,a}) = \hat{\sigma}_{GM}^2 \widehat{M}_{GM}^{-1} \hat{Q}_{GM,gw} \widehat{M}_{GM}^{-1} \ , \tag{1.101}$$

where

$$\hat{Q}_{GM,gw} = \sum_{i=1}^{n} \frac{\eta^2(\mathbf{x}_i, r_i^{GM}/\hat{\sigma}_{GM})}{1 - w_i^{GM}} \mathbf{x}_i \mathbf{x}_i^T \ . \tag{1.102}$$

1.3.3 Formulas summary for chapter one

The ordinary least squares estimate

Name	Formula for, or description of $\hat{\beta}_{(i)}$	Formula for pseudovalues $\hat{\beta}^i$	Formula for variance of $\hat{\beta}$
ordinary jackknife	$\hat{\beta} - \frac{(X^TX)^{-1}\mathbf{x}_i r_i}{1-h_i}$	$n\hat{\beta} - (n-1)\hat{\beta}_{(i)}$	$\frac{1}{n(n-1)}\sum(\hat{\beta}^i - \hat{\beta}_J)(\hat{\beta}^i - \hat{\beta}_J)^T$
§1.2.3.1	(1.22)	(1.18)	(1.20)
weighted jackknife	same as above	$\hat{\beta} + n(1-h_i)(\hat{\beta} - \hat{\beta}_{(i)})$	$\frac{n}{n-p}(X^TX)^{-1}(\sum r_i^2 \mathbf{X}_i\mathbf{X}_i^T)(X^TX)^{-1}$
§1.2.3.2		(1.21)	(1.24)
general weighted	same as above	$\hat{\beta} + (n-p)^{1/2}(\hat{\beta}_{(i)} - \hat{\beta})$	$(X^TX)^{-1}(\sum \frac{r_i^2}{1-h_i}\mathbf{X}_i\mathbf{X}_i^T)(X^TX)^{-1}$
jackknife §1.2.3.3		(1.25)	(1.30)

The robust estimate in the general function form

Name	Formula for, or description of $\hat{\beta}_{\phi(i)}$	Formula for pseudovalues $\hat{\beta}^i_\phi$	Formula for variance of $\hat{\beta}$
OJR	$\sum_{j\neq i}\phi(z_j, \hat{\beta}_{\phi(i)}, \hat{\sigma}_\phi) = 0$	$n\hat{\beta}_\phi - (n-1)\hat{\beta}_{\phi(i)}$	$\frac{1}{n(n-1)}\sum(\hat{\beta}^i_\phi - \hat{\beta}^\phi_J)(\hat{\beta}^i_\phi - \hat{\beta}^\phi_J)^T$
§1.3.1.1	(1.33)	(1.34)	(1.36)
WJR	same as above	$\hat{\beta}_\phi + n(1-w_i^\phi)(\hat{\beta}_\phi - \hat{\beta}_{\phi(i)})$	$\frac{1}{n(n-p)}\sum(\hat{\beta}^i_{\phi,w} - \hat{\beta}^\phi_{J,w})(\hat{\beta}^i_{\phi,w} - \hat{\beta}^\phi_{J,w})^T$
§1.3.1.2		(1.45)	(1.47)
GWJR	same as above	$\hat{\beta}_\phi + (n-p)^{1/2}(\hat{\beta}_{\phi(i)} - \hat{\beta}_\phi)$	$\sum\alpha_i^\phi(\hat{\beta}^i_{\phi,gw} - \hat{\beta}^\phi_{J,gw})(\hat{\beta}^i_{\phi,gw} - \hat{\beta}^\phi_{J,gw})^T$
§1.3.1.3		(1.59)	(1.61)

The M-estimate (For the exact procedure)

Name	Formula for, or description of $\hat{\beta}_{M(i)}$	Formula for pseudovalues $\hat{\beta}^i_M$	Formula for variance of $\hat{\beta}$
OJM	$\sum_{j\neq i}\psi(r^M_{j(i)}/\hat{\sigma}_M)\mathbf{x}_j = 0$	$n\hat{\beta}_M - (n-1)\hat{\beta}_{M(i)}$	$\frac{1}{n(n-1)}\sum(\hat{\beta}^i_M - \hat{\beta}^M_J)(\hat{\beta}^i_M - \hat{\beta}^M_J)^T$
§1.3.1.1	(1.38)	(1.37)	(1.40)
WJM	same as above	$\hat{\beta}_M + n(1-w_i^M)(\hat{\beta}_M - \hat{\beta}_{M(i)})$	$\frac{1}{n(n-p)}\sum(\hat{\beta}^i_{M,w} - \hat{\beta}^M_{J,w})(\hat{\beta}^i_{M,w} - \hat{\beta}^M_{J,w})^T$
§1.3.1.2		(1.51)	(1.53)
GWJM	same as above	$\hat{\beta}_M + (n-p)^{1/2}(\hat{\beta}_{M(i)} - \hat{\beta}_M)$	$\sum\alpha_i^M(\hat{\beta}^i_{M,gw} - \hat{\beta}^M_{J,gw})(\hat{\beta}^i_{M,gw} - \hat{\beta}^M_{J,gw})^T$
§1.3.1.3		(1.62)	(1.64)

The GM-estimate (For the exact procedure)

Name	Formula for, or description of $\hat{\beta}_{GM(i)}$	Formula for pseudovalues $\hat{\beta}^{i}_{GM}$	Formula for variance of $\hat{\beta}$
OJGM	$\sum \eta(\mathbf{x}_j, \frac{r^{GM}_{j(i)}}{\hat{\sigma}_{GM}})\mathbf{x}_j = 0$	$n\hat{\beta}_{GM} - (n-1)\hat{\beta}_{GM(i)}$	$\frac{1}{n(n-1)} \sum(\hat{\beta}^{i}_{GM} - \hat{\beta}^{GM}_{J})(\hat{\beta}^{i}_{GM} - \hat{\beta}^{GM}_{J})^T$
§1.3.1.1	(1.42)	(1.41)	(1.43)
WJGM	same as above	$\hat{\beta}_{GM} + n(1-w^{GM}_{i})(\hat{\beta}_{GM} - \hat{\beta}_{GM(i)})$	$\frac{1}{n(n-p)} \sum(\hat{\beta}^{i}_{GM,w} - \hat{\beta}^{GM}_{J,w})(\hat{\beta}^{i}_{GM,w} - \hat{\beta}^{GM}_{J,w})^T$
§1.3.1.2		(1.56)	(1.58)
GWJGM	same as above	$\hat{\beta}_{GM} + (n-p)^{1/2}(\hat{\beta}_{GM(i)} - \hat{\beta}_{GM})$	$\sum \alpha^{GM}_{i}(\hat{\beta}^{i}_{GM,gw} - \hat{\beta}^{GM}_{J,gw})(\hat{\beta}^{i}_{GM,gw} - \hat{\beta}^{GM}_{J,gw})^T$
§1.3.1.3		(1.65)	(1.67)

The M-estimate (For the approximate procedure)

Name	Formula for, or description of $\hat{\beta}^{a}_{M(i)}$	Formula for pseudovalues $\hat{\beta}^{i,a}_{M}$	Formula for variance of $\hat{\beta}$
OJMA	$\hat{\beta}_M - \hat{\sigma}_M \frac{\psi(\frac{r^{M}_{i}}{\hat{\sigma}_M})}{1-w^{M}_{i}} \widehat{M}^{-1}_{M}\mathbf{x}_i$	$\hat{\beta}_M + (n-1)(\hat{\beta}_M - \hat{\beta}^{a}_{M(i)})$	$\frac{1}{n(n-1)} \sum(\hat{\beta}^{i,a}_{M} - \hat{\beta}^{M,a}_{J})(\hat{\beta}^{i,a}_{M} - \hat{\beta}^{M,a}_{J})^T$
§1.3.2.3 (1)	(1.73)	(1.78)	(1.80)
WJMA	same as above	$\hat{\beta}_M + n(1-w^{M}_{i})(\hat{\beta}_M - \hat{\beta}^{a}_{M(i)})$	$\frac{n}{n-p}\hat{\sigma}^2_M \widehat{M}^{-1}_{M}\widehat{Q}_M \widehat{M}^{-1}_{M}$
§1.3.2.3 (2)		(1.87)	(1.89)
GWJMA	same as above	$\hat{\beta}_M + (n-p)^{1/2}(\hat{\beta}^{a}_{M(i)} - \hat{\beta}_M)$	$\hat{\sigma}^2_M \widehat{M}^{-1}_{M}\widehat{Q}_{M,gw}\widehat{M}^{-1}_{M}$
§1.3.2.3 (3)		(1.97)	(1.98)

The GM-estimate (For the approximate procedure)

Name	Formula for, or description of $\hat{\beta}^{a}_{GM(i)}$	Formula for pseudovalues $\hat{\beta}^{i,a}_{GM}$	Formula for variance of $\hat{\beta}$
OJGMA	$\hat{\beta}_{GM} - \hat{\sigma}_{GM} \frac{\eta(\mathbf{x}_i, \frac{r^{GM}_{i}}{\hat{\sigma}_{GM}})}{1-w^{GM}_{i}} \widehat{M}^{-1}_{GM}\mathbf{x}_i$	$\hat{\beta}_{GM} + (n-1)(\hat{\beta}_{GM} - \hat{\beta}^{a}_{GM(i)})$	$\sum \frac{(\hat{\beta}^{i,a}_{GM} - \hat{\beta}^{GM,a}_{J})(\hat{\beta}^{i,a}_{GM} - \hat{\beta}^{GM,a}_{J})^T}{n(n-1)}$
§1.3.2.3 (1)	(1.74)	(1.81)	(1.83)
WJGMA	same as above	$\hat{\beta}_{GM} + n(1-w^{GM}_{i})(\hat{\beta}_{GM} - \hat{\beta}^{a}_{GM(i)})$	$\frac{n}{n-p}\hat{\sigma}^2_{GM} \widehat{M}^{-1}_{GM}\widehat{Q}_{GM} \widehat{M}^{-1}_{GM}$
§1.3.2.3 (2)		(1.91)	(1.92)
GWJGMA	same as above	$\hat{\beta}_{GM} + (n-p)^{1/2}(\hat{\beta}^{a}_{GM(i)} - \hat{\beta}_{GM})$	$\hat{\sigma}^2_{GM} \widehat{M}^{-1}_{GM}\widehat{Q}_{GM,gw}\widehat{M}^{-1}_{GM}$
§1.3.2.3 (3)		(1.100)	(1.101)

2 Simulation Study

2.1 The Model and The Data

In doing our simulation, we consider the following regression model

$$y_i = \beta_0 + \beta_1 \mathbf{x}_i + \varepsilon_i \qquad i = 1, 2, \ldots, n \tag{2.1}$$

where β_0 and β_1 are parameters. In our simulation, we set $\beta_0 = 1$ and $\beta_1 = 1$.

We choose the error terms ε_i $(i = 1, 2, \ldots, n)$ such that some of ε_i are distributed according to the normal distribution with mean zero and variance equal to one, and the rest of the ε_i follow a more heavy tailed distribution. The ratio of these two parts are $(1 - \nu)/\nu$, where ν satisfies $0 < \nu < 1$.

Based on the above idea, the random variable ε_i can be generated by the formula

$$\varepsilon_i \sim (1 - \delta_i) \cdot \Phi(t) + \delta_i \cdot F(t) \,, \tag{2.2}$$

where $\Phi(t)$ is the standard normal distribution function $N(0, 1)$ and $F(s)$ is the t-distribution on four degrees of freedom, and the variable δ_i $(i = 1, 2, \ldots, n)$ satisfies

$$\delta_i = \begin{cases} 0 & \text{with probability } 1 - \nu, \\ 1 & \text{with probability } \nu. \end{cases} \tag{2.3}$$

It is easy to see that the error terms ε_i , do not satisfy the classical regression assumptions when $\nu \neq 0$.

Now we consider constructing the design matrix X in our simulation study. The form of design matrix is $X = (\mathbf{1}, \mathbf{x})$, where $\mathbf{1}$ is a vector with all elements equal to 1. Similar to the way we generate the error terms ε_i, we define the elements of the vector $\mathbf{x}$ by

$$\mathbf{x}_i = (1 - \frac{\delta_i}{2}) \cdot N(-1,1) + \frac{\delta_i}{2}|t_2| \ , \qquad i = 1, \ldots, n \tag{2.4}$$

where $N(-1,1)$ is normal distribution with mean equals to -1 and the variance equals to 1. The value of t_2 represents a t-distribution on two degree of freedom.

The response variable $\mathbf{y}$ is generated in such a way that $(1 - \nu/2) \cdot 100$ percent of the data are formed by

$$\mathbf{y}_i = \beta_0 + \beta_1 \cdot \mathbf{x}_i + \varepsilon_i \ , \tag{2.5}$$

and the rest of $\mathbf{y}_i$ values are formed by

$$\mathbf{y}_i = \beta_0 + 0.1\beta_1 \cdot \mathbf{x}_i + \varepsilon_i \ . \tag{2.6}$$

The way to generate the paired data $(\mathbf{y}_i, \mathbf{x}_i)$ is as follow : $(1 - \nu) \cdot 100$ percent of the $(\mathbf{y}_i, \mathbf{x}_i)$ follow the model $\mathbf{y}_i = \beta_0 + \beta_1 \cdot \mathbf{x}_i + \varepsilon_i$, where $\varepsilon_i \sim N(0,1)$ and $\mathbf{x}_i \sim N(-1,1)$. $\nu/2 \cdot 100$ percent of the $(\mathbf{y}_i, \mathbf{x}_i)$ follow the same model but $\varepsilon_i \sim N(0,1)$ and $\mathbf{x}_i \sim |t_2|$. The rest of $\nu/2 \cdot 100$ percent of the $(\mathbf{y}_i, \mathbf{x}_i)$ follow the model $\mathbf{y}_i = \beta_0 + 0.1\beta_1 \cdot \mathbf{x}_i + \varepsilon_i$, where $\varepsilon_i \sim t_4$ and $\mathbf{x}_i \sim |t_2|$. The random variables $\varepsilon_1, \ldots, \varepsilon_n$ and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent.

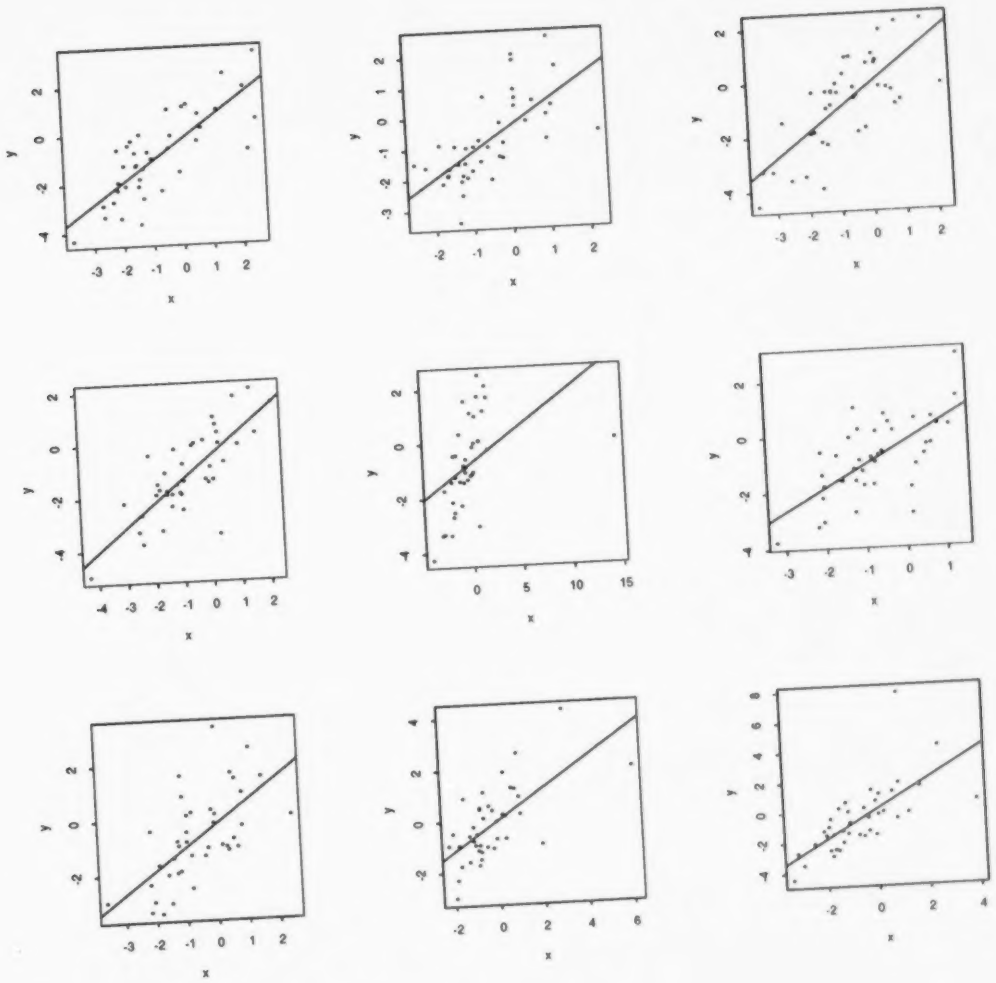


Figure 1: Scatterplot of sample data generated by taking $\nu = 0.2$, with least squares regression lines

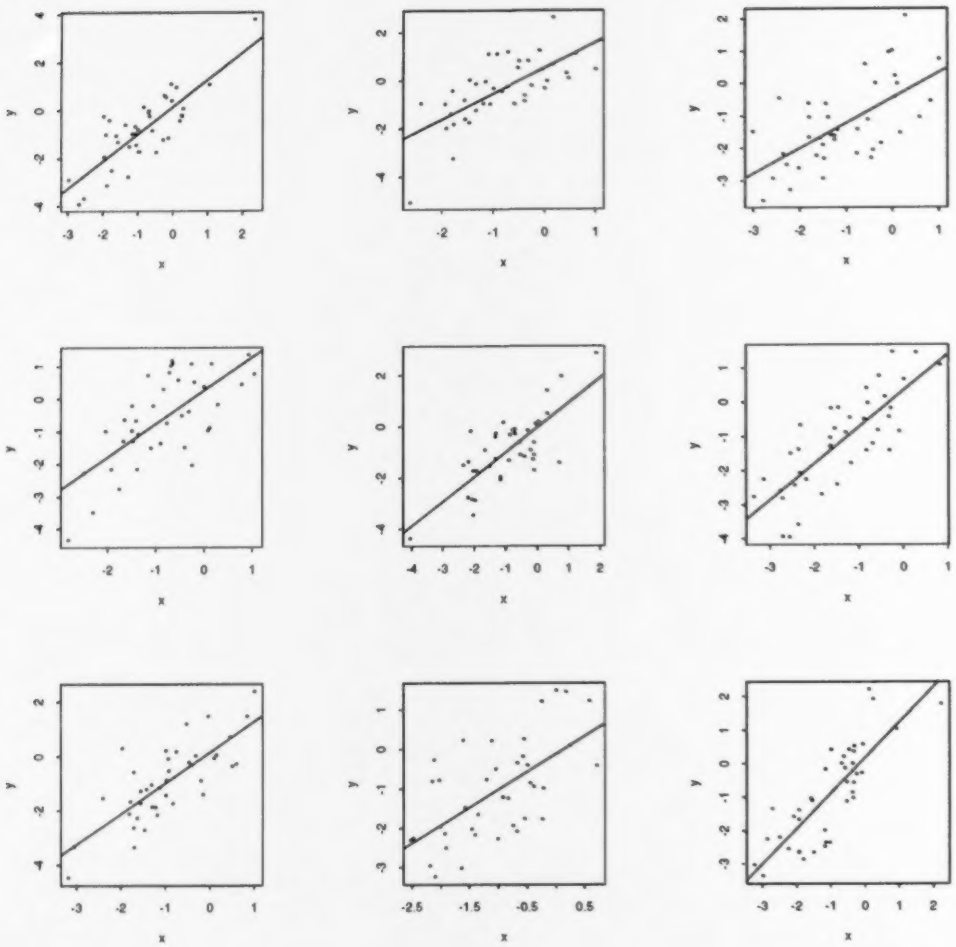


Figure 2: Scatterplot of sample data generated by taking $\nu = 0$, with least squares regression lines

Based on how we generate the design matrix and the response variable, we know that when $\nu \neq 0$ there are outliers or possible influential points in the design matrix X and the response variable $\mathbf{y}$. From Figure 1 and Figure 2, we can have a general idea about the structure of those variables.

2.2 Procedures

We consider the cases of $\nu = 0$, 0.2 and $n = 20$, 40 in our simulation study. The procedures we used in our the simulation run are:

- Ordinary least squares estimate (LS)
- Huber's M-estimate (M.H)
- Mallows' generalized M-estimate (GM.M)
- Schweppe's generalized M-estimate (GM.S)

When doing M-estimation, we choose Huber's ψ function with $k = 1$. For the GM-estimate, we use Mallows' type η function. i.e.

$$\eta(x, \frac{r}{\sigma}) = w(x) \cdot \psi(\frac{r}{\sigma}) \tag{2.7}$$

and Schweppe's type η function. i.e.

$$\eta(x, \frac{r}{\sigma}) = w(x) \cdot \psi(\frac{r}{\sigma} \cdot v(x)) \tag{2.8}$$

where $w(x) = \sqrt{1-h_i}$, and $v(x) = 1/w(x)$, h_i are the diagonal elements of hat matrix H and the ψ function is the same function as is being used in the M-estimate case.

More covariance estimations have been introduced when we conduct the simulation study. Those estimates are basically obtained by different combinations of $\widehat{M}$ and $\hat{Q}$ matrices which are defined as following.

(1). Exchangeable $\widehat{M}$

The exchangeable $\widehat{M}$ are only have values for Huber and Mallows cases. (The exchangeable implies that $\mathbf{x}$ and ε are independent, e.g. $E[\psi'(\varepsilon)\mathbf{x}\mathbf{x}^T] = E\{\psi'(\varepsilon)\}E[\mathbf{x}\mathbf{x}^T]$). For Huber's M-estimate, the exchangeable $\widehat{M}$ matrix is

$$\widehat{M}^{ex}_{M,h} = \frac{1}{n}\sum_{i=1}^{n}\psi'(\frac{r_i^{M,h}}{\hat{\sigma}_{M,h}})\cdot\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T \ , \tag{2.9}$$

and for Mallows' GM-estimate, we have

$$\widehat{M}^{ex}_{GM,m} = \frac{1}{n}\sum_{i=1}^{n}\psi'(\frac{r_i^{GM,m}}{\hat{\sigma}_{GM,m}})\cdot\sum_{i=1}^{n}w(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T \ . \tag{2.10}$$

(2). Non-exchangeable $\widehat{M}$

For Huber's M-estimate , non-exchangeable $\widehat{M}$ is the same as formula (1.50). For Mallows' and Schweppe's GM-estimate, we can get the corresponding formulas from formula (1.55). For Mallows' GM-estimate, we have

$$\widehat{M}^{non}_{GM,m} = \sum_{i=1}^{n}w(\mathbf{x}_i)\psi'(\frac{r_i^{GM,m}}{\hat{\sigma}_{GM,m}})\mathbf{x}_i\mathbf{x}_i^T \ , \tag{2.11}$$

and for Schweppe's GM-estimate, the expression is

$$\widehat{M}^{non}_{GM,s} = \sum_{i=1}^{n} \psi'(\frac{r_i^{GM,s}}{\hat{\sigma}_{GM,s} \cdot w(\mathbf{x}_i)}) \mathbf{x}_i \mathbf{x}_i^T \ . \tag{2.12}$$

(3). Markatou, Stahel and Ronchetti's type $\widehat{M}$

Markatou, Stahel and Ronchetti (1991) gave following form of $\widehat{M}$,

$$\widehat{M}^{MSR} = \sum_{i=1}^{n} \eta^2(\mathbf{x}_i, \frac{r_i}{\hat{\sigma}}) \cdot (\frac{\sum_{j=1}^{n} \eta'(\mathbf{x}_i, r_j/\hat{\sigma})/n}{\sum_{j=1}^{n} \eta^2(\mathbf{x}_i, r_j/\hat{\sigma})/(n-p)}) \mathbf{x}_i \mathbf{x}_i^T \ . \tag{2.13}$$

For Huber's M-estimate case, it can be simplified as

$$\widehat{M}^{MSR}_{M,h} = \frac{\sum_{i=1}^{n} \psi'(r_i^{M,h}/\hat{\sigma}_{M,h})/n}{\sum_{i=1}^{n} \psi^2(r_i^{M,h}/\hat{\sigma}_{M,h})/(n-p)} \sum_{i=1}^{n} \psi^2(r_i^{M,h}/\hat{\sigma}_{M,h}) \mathbf{x}_i \mathbf{x}_i^T \ , \tag{2.14}$$

and for Mallows' GM-estimate case, it has the simplification form

$$\widehat{M}^{MSR}_{GM,m} = \frac{\sum_{i=1}^{n} \psi'(r_i^{GM,m}/\hat{\sigma}_{GM,m})/n}{\sum_{i=1}^{n} \psi^2(r_i^{GM,m}/\hat{\sigma}_{GM,m})/(n-p)} \sum_{i=1}^{n} w(\mathbf{x}_i) \psi^2(r_i^{GM,m}/\hat{\sigma}_{GM,m}) \mathbf{x}_i \mathbf{x}_i^T \ . \tag{2.15}$$

(4). Exchangeable $\hat{Q}$

We have the exchangeable $\hat{Q}$ matrix for Huber's and Mallows' estimates. For Huber's M-estimate case, it is

$$\hat{Q}^{ex}_{M,h} = \frac{1}{n-p} \sum_{i=1}^{n} \psi^2(\frac{r_i^{M,h}}{\hat{\sigma}_{M,h}}) \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \ , \tag{2.16}$$

for Mallows' type GM-estimate case, we have

$$\hat{Q}^{ex}_{GM,m} = \frac{1}{n-p} \sum_{i=1}^{n} \psi^2(\frac{r_i^{GM,m}}{\hat{\sigma}_{GM,m}}) \sum_{i=1}^{n} w^2(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^T \ . \tag{2.17}$$

(5). Non-exchangeable $\hat{Q}$

We can find that the formula (1.90) is the non-exchangeable $\hat{Q}$ for Huber's M-estimate case. Also, we can get the expressions for Mallows' and Schweppe's GM-estimate cases by using formula (1.93). They are

$$\hat{Q}^{non}_{GM,m} = \sum_{i=1}^{n} w^2(\mathbf{x}_i)\psi^2(\frac{r_i^{GM,m}}{\hat{\sigma}_{GM,m}})\mathbf{x}_i\mathbf{x}_i^T \ , \tag{2.18}$$

and

$$\hat{Q}^{non}_{GM,s} = \sum_{i=1}^{n} w^2(\mathbf{x}_i)\psi^2(\frac{r_i^{GM,s}}{\hat{\sigma}_{GM,s}\cdot w(\mathbf{x}_i)})\mathbf{x}_i\mathbf{x}_i^T \ . \tag{2.19}$$

(6). Jackknife non-exchangeable $\hat{Q}$

We define the covariance estimate for the general weighted approximate jackknife as jackknife non-exchangeable $\hat{Q}$ which have the expression as (1.99) for Huber's M-estimate case. For Mallows' and Schweppe's GM-estimates, we can get corresponding expressions from formula (1.102). For Mallows' type GM-estimate, the jackknife non-exchangeable $\hat{Q}$ is

$$\hat{Q}^{J,non}_{GM,m} = \sum_{i=1}^{n} \frac{w^2(\mathbf{x}_i)\psi^2(r_i^{GM,m}/\hat{\sigma}_{GM,m})\mathbf{x}_i\mathbf{x}_i^T}{1-w_i^{GM,m}} \ , \tag{2.20}$$

and for Schweppe's type GM-estimate, the jackknife non-exchangeable $\hat{Q}$ is

$$\hat{Q}^{J,non}_{GM,s} = \sum_{i=1}^{n} \frac{w^2(\mathbf{x}_i)\psi^2(\frac{r_i^{GM,s}}{\hat{\sigma}_{GM,s}\cdot w(\mathbf{x}_i)})\mathbf{x}_i\mathbf{x}_i^T}{1-w_i^{GM,s}} \ . \tag{2.21}$$

(7). Jackknife exchangeable $\hat{Q}$

For jackknife exchangeable $\hat{Q}$, we have expressions for Huber's and Mallows estimates. The Huber's jackknife exchangeable $\hat{Q}$ is

$$\hat{Q}^{J,ex}_{M,h} = \frac{1}{n-p}\sum_{i=1}^{n}\psi^2(\frac{r_i^{M,h}}{\hat{\sigma}_{M,h}})\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^T}{1-w_i^{M,h}} , \tag{2.22}$$

and the Mallows' jackknife exchangeable $\hat{Q}$ is

$$\hat{Q}^{J,ex}_{GM,m} = \frac{1}{n-p}\sum_{i=1}^{n}\psi^2(\frac{r_i^{GM,m}}{\hat{\sigma}_{GM,m}})\sum_{i=1}^{n}\frac{w^2(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T}{1-w_i^{GM,m}} \tag{2.23}$$

(8). Ordinary jackknife $\hat{Q}$

We define the ordinary jackknife $\hat{Q}$ as

$$\hat{Q}^{J,ord} = \sum_{i=1}^{n}\frac{\eta^2(\mathbf{x}_i, r_i/\hat{\sigma})}{(1-w_i)^2}\mathbf{x}_i\mathbf{x}_i^T - \frac{1}{n}(\sum_{i=1}^{n}\frac{\eta(r_i/\hat{\sigma})\mathbf{x}_i}{1-w_i})(\sum_{i=1}^{n}\frac{\eta(r_i/\hat{\sigma})\mathbf{x}_i}{1-w_i})^T . \tag{2.24}$$

From above formula, we can obtain the ordinary jackknife $\hat{Q}$ for Huber, Mallows and Schweppe's cases.

Based on above $\widehat{M}$ and $\hat{Q}$ matrices, we can form our covariance estimates for Huber's, Mallows' and Schweppe's type robust estimation. Those estimates are :

Huber's case

(1). $V_{h.1.a} = \hat{\sigma}^2_{M,h}\cdot\widehat{M}^{non-1}_{M,h}\cdot\hat{Q}^{non}_{M,h}\cdot\widehat{M}^{non-1}_{M,h}$

(2). $V_{h.1.b} = \hat{\sigma}^2_{M,h}\cdot\widehat{M}^{ex-1}_{M,h}\cdot\hat{Q}^{ex}_{M,h}\cdot\widehat{M}^{ex-1}_{M,h}$

(3). $V_{h.1.c} = \hat{\kappa}^2\cdot V_{h.1.b}$

(4). $V_{h.1.d} = \hat{\sigma}^2_{M,h} \cdot \widehat{M}^{non-1}_{M,h} \cdot \hat{Q}^{ex}_{M,h} \cdot \widehat{M}^{non-1}_{M,h}$

(5). $V_{h.1.e} = \hat{\kappa}^{-1} \cdot V_{h.1.d}$

(6). $V_{h.2.a} = \hat{\sigma}^2_{M,h} \cdot \widehat{M}^{MSR-1}_{M,h} \cdot \hat{Q}^{non}_{M,h} \cdot \widehat{M}^{MSR-1}_{M,h}$

(7). $V_{h.3.a}$ = exact ordinary jackknife estimate for Huber's case

(8). $V_{h.3.b}$ = exact weighted jackknife estimate for Huber's case

(9). $V_{h.3.c}$ = exact general weighted jackknife estimate for Huber's case

(10). $V_{h.4.a} = \frac{n-1}{n}\hat{\sigma}^2_{M,h} \cdot \widehat{M}^{non-1}_{M,h} \cdot \hat{Q}^{J,ord}_{M,h} \cdot \widehat{M}^{non-1}_{M,h}$

(approximate ordinary jackknife estimate for Huber's case)

(11). $V_{h.5.a} = \frac{n}{n-p} \cdot V_{h.1.a}$

(approximate weighted jackknife estimate for Huber's case)

(12). $V_{h.5.b} = \frac{n}{n-p} \cdot V_{h.1.b}$

(13). $V_{h.5.c} = \frac{n}{n-p} \cdot V_{h.1.c}$

(14). $V_{h.5.d} = \frac{n}{n-p} \cdot V_{h.1.d}$

(15). $V_{h.5.e} = \frac{n}{n-p} \cdot V_{h.1.e}$

(16). $V_{h.6.a} = \hat{\sigma}^2_{M,h} \cdot \widehat{M}^{non-1}_{M,h} \hat{Q}^{J,non}_{M,h} \cdot \widehat{M}^{non-1}_{M,h}$

(approximate general weighted jackknife estimate for Huber's case)

(17). $V_{h.6.b} = \hat{\sigma}^2_{M,h} \cdot \widehat{M}^{ex-1}_{M,h} \hat{Q}^{J,non}_{M,h} \cdot \widehat{M}^{ex-1}_{M,h}$

(18). $V_{h.6.c} = \hat{\sigma}^2_{M,h} \cdot \widehat{M}^{non-1}_{M,h} \hat{Q}^{J,ex}_{M,h} \cdot \widehat{M}^{non-1}_{M,h}$

(19). $V_{h.6.d} = \hat{\sigma}^2_{M,h} \cdot \widehat{M}^{ex-1}_{M,h} \hat{Q}^{J,ex}_{M,h} \cdot \widehat{M}^{ex-1}_{M,h}$

(20). $V_{h.7.a} = \hat{\kappa} \cdot \hat{\sigma}^2_{M,h} \frac{\sum \psi^2(r_i^{M,h}/\hat{\sigma}_{M,h})/(n-p)}{\sum \psi'(r_i^{M,h}/\hat{\sigma}_{M,h})/n} \cdot \widehat{M}^{non-1}_{M,h}$

Mallows' case

(1). $V_{m.1.a} = \hat{\sigma}^2_{GM,m} \cdot \widehat{M}^{non-1}_{GM,m} \cdot \hat{Q}^{non}_{GM,m} \cdot \widehat{M}^{non-1}_{GM,m}$

(2). $V_{m.1.b} = \hat{\sigma}^2_{GM,m} \cdot \widehat{M}^{ex-1}_{GM,m} \cdot \hat{Q}^{ex}_{GM,m} \cdot \widehat{M}^{ex-1}_{GM,m}$

(3). $V_{m.1.c} = \hat{\kappa}^2 \cdot V_{m.1.b}$

(4). $V_{m.1.d} = \hat{\sigma}^2_{GM,m} \cdot \widehat{M}^{non-1}_{GM,m} \cdot \hat{Q}^{ex}_{GM,m} \cdot \widehat{M}^{non-1}_{GM,m}$

(5). $V_{m.1.e} = \hat{\kappa}^{-1} \cdot V_{m.1.d}$

(6). $V_{m.2.a} = \hat{\sigma}^2_{GM,m} \cdot \widehat{M}^{MSR-1}_{GM,m} \cdot \hat{Q}^{non}_{GM,m} \cdot \widehat{M}^{MSR-1}_{GM,m}$

(7). $V_{m.3.a}$ = exact ordinary jackknife estimate for Mallows' case

(8). $V_{m.3.b}$ = exact weighted jackknife estimate for Mallows' case

(9). $V_{m.3.c}$ = exact general weighted jackknife estimate for Mallows' case

(10). $V_{m.4.a} = \frac{n-1}{n}\hat{\sigma}^2_{GM,m} \cdot \widehat{M}^{non-1}_{GM,m} \cdot \hat{Q}^{J,ord}_{GM,m} \cdot \widehat{M}^{nonx-1}_{GM,m}$

(approximate ordinary jackknife estimate for Mallows' case)

(11). $V_{m.5.a} = \frac{n}{n-p} \cdot V_{m.1.a}$

(approximate weighted jackknife estimate for Mallows' case)

(12). $V_{m.5.b} = \frac{n}{n-p} \cdot V_{m.1.b}$

(13). $V_{m.5.c} = \frac{n}{n-p} \cdot V_{m.1.c}$

(14). $V_{m.5.d} = \frac{n}{n-p} \cdot V_{m.1.d}$

(15). $V_{m.5.e} = \frac{n}{n-p} \cdot V_{m.1.e}$

(16). $V_{m.6.a} = \hat{\sigma}^2_{GM,m} \cdot \widehat{M}^{non-1}_{GM,m}\hat{Q}^{J,non}_{GM,m} \cdot \widehat{M}^{non-1}_{GM,m}$

(approximate general weighted jackknife estimate for Mallows' case)

(17). $V_{m.6.b} = \hat{\sigma}^2_{GM,m} \cdot \widehat{M}^{ex-1}_{GM,m}\hat{Q}^{J,non}_{GM,m} \cdot \widehat{M}^{ex-1}_{GM,m}$

(18). $V_{m.6.c} = \hat{\sigma}^2_{GM,m} \cdot \widehat{M}^{non-1}_{GM,m}\hat{Q}^{J,ex}_{GM,m} \cdot \widehat{M}^{non-1}_{GM,m}$

(19). $V_{m.6.d} = \hat{\sigma}^2_{GM,m} \cdot \widehat{M}^{ex-1}_{GM,m}\hat{Q}^{J,ex}_{GM,m} \cdot \widehat{M}^{ex-1}_{GM,m}$

Schweppe's case

(1). $V_{s.1.a} = \hat{\sigma}^2_{GM,s} \cdot \widehat{M}^{non-1}_{GM,s} \cdot \widehat{Q}^{non}_{GM,s} \cdot \widehat{M}^{non-1}_{GM,s}$

(2). $V_{s.2.a} = \hat{\sigma}^2_{GM,s} \cdot \widehat{M}^{MSR-1}_{GM,s} \cdot \widehat{Q}^{non}_{GM,s} \cdot \widehat{M}^{MSR-1}_{GM,s}$

(3). $V_{s.3.a}$ = exact ordinary jackknife estimate for Schweppe's case

(4). $V_{s.3.b}$ = exact weighted jackknife estimate for Schweppe's case

(5). $V_{s.3.c}$ = exact general weighted jackknife estimate for Schweppe's case

(6). $V_{s.4.a} = \frac{n-1}{n}\hat{\sigma}^2_{GM,s} \cdot \widehat{M}^{ex-1}_{GM,s} \cdot \widehat{Q}^{J,ord}_{GM,s} \cdot \widehat{M}^{ex-1}_{GM,s}$

(approximate ordinary jackknife estimate for Schweppe's case)

(7). $V_{s.5.a} = \frac{n}{n-p} \cdot V_{s.1.a}$

(approximate weighted jackknife estimate for Schweppe's case)

(8). $V_{s.6.a} = \hat{\sigma}^2_{GM,s} \cdot \widehat{M}^{non-1}_{GM,s} \widehat{Q}^{J,non}_{GM,s} \cdot \widehat{M}^{non-1}_{GM,s}$

(approximate general weighted jackknife estimate for Schweppe's case)

Note:

In above formulas, the value of $\hat{\kappa}$ is an estimate of

$$\kappa = 1 + \frac{p}{n}\frac{Var(\psi')}{(E\psi')^2}$$

2.3 The Measurement

In our simulation study, we considered several statistics in measuring our procedures. The bias of the estimate was applied to measure the accuracy of the estimates.

$$bias = \hat{\beta}_i - \beta_i \qquad i = 0,\ 1 \tag{2.25}$$

The relative bias of the variance estimation is applied to measure the efficiency of those covariance estimates. The relative bias of the variance estimation is defined as

$$\gamma_{ij} = \frac{\bar{v}_{ij}^{E} - \bar{v}_{ij}^{S}}{|\bar{v}_{ij}^{S}|} \cdot 100 \ , \tag{2.26}$$

where

$\bar{v}_{ij}^{E}$ and $\bar{v}_{ij}^{S}$ are the ith row and jth column element of matrices $\bar{V}^{E}$ and $\bar{V}^{S}$ respectively;

$\bar{V}^{E}$ = average of the covariance estimate for total simulation run;

$\bar{V}^{S}$ = variance of the estimator of parameter for total simulation run.

2.4 The Simulation Results

All of the following simulation results are obtained based on 1000 runs.

2.4.1 The Bias of the Estimates

The bias of the estimates can be found in Table 1. From the results, we can see that for both n equals 20 and 40, when the contamination ratio $\nu = 0$, the least squares estimates (the original LS, and the three types of jackknife estimates) performed well. Generally, they have smaller biases than the robust procedures. But when the data is contaminated (in our case the contamination ratio $\nu = 0.2$), those three types of robust procedures have smaller biases than the least squares procedures. Within each robust procedure, we can find that the Mallows' and Schweppe's type GM-estimates have smaller biases than Huber's M-estimate for both $n = 20$ and $n = 40$ cases. We

Type of estimate	$n = 20$				$n = 40$			
	$\nu = 0$		$\nu = 0.2$		$\nu = 0$		$\nu = 0.2$	
	bias β_0	bias β_1	bias β_0	bias β_1	bias β_0	bias β_1	bias β_0	bias β_1
LS	0.00425	−0.00290	−0.17985	−0.16532	0.01084	0.00448	−0.21136	−0.19414
LSOJ	0.00531	−0.00275	−0.21995	−0.20415	0.01064	0.00420	−0.25014	−0.23777
LSWJ	0.00425	−0.00290	−0.17985	−0.16532	0.01084	0.00448	−0.21136	−0.19414
LSGWJ	0.00425	−0.00290	−0.17985	−0.16532	0.01084	0.00448	−0.21136	−0.19414
H	0.00527	−0.00245	−0.15709	−0.14481	0.01274	0.00618	−0.16928	−0.15042
HOJ	0.00165	−0.00132	−0.16518	−0.15176	0.01450	0.00741	−0.17339	−0.15427
HWJ	0.00246	−0.00064	−0.15091	−0.13715	0.01325	0.00680	−0.16616	−0.14645
HGWJ	0.00594	−0.00288	−0.15855	−0.14661	0.01266	0.00608	−0.16978	−0.15106
HOJA	0.00582	−0.00176	−0.17906	−0.16972	0.01374	0.00679	−0.18063	−0.16364
HWJA	0.00531	−0.00243	−0.15717	−0.14486	0.01268	0.00617	−0.16944	−0.15058
HGWJA	0.00527	−0.00245	−0.15708	−0.14479	0.01275	0.00618	−0.16925	−0.15039
M	0.00514	−0.00214	−0.13542	−0.12307	0.01286	0.00630	−0.14994	−0.12829
MOJ	0.00372	0.00072	−0.14935	−0.13412	0.01369	0.00662	−0.16127	−0.13715
MWJ	0.00422	0.00097	−0.14528	−0.13254	0.01247	0.00598	−0.16015	−0.13767
MGWJ	0.00535	−0.00287	−0.13309	−0.12084	0.01293	0.00635	−0.14828	−0.12677
MOJA	0.00518	−0.00190	−0.13735	−0.12240	0.01388	0.00698	−0.15056	−0.12739
MWJA	0.00519	−0.00212	−0.13536	−0.12298	0.01281	0.00630	−0.14993	−0.12826
MGWJA	0.00512	−0.00215	−0.13543	−0.12309	0.01287	0.00630	−0.14994	−0.12830
S	0.00528	−0.00217	−0.13572	−0.12261	0.01316	0.00641	−0.15017	−0.12803
SOJ	0.00246	0.00291	−0.15056	−0.13087	0.01682	0.00864	−0.16366	−0.13859
SWJ	0.00106	0.00256	−0.14768	−0.13119	0.01592	0.00838	−0.16259	−0.13916
SGWJ	0.00628	−0.00328	−0.13291	−0.12059	0.01272	0.00610	−0.14815	−0.12623
SOJA	0.00746	−0.00155	−0.13637	−0.12007	0.01390	0.00674	−0.15089	−0.12719
SWJA	0.00534	−0.00214	−0.13563	−0.12249	0.01312	0.00642	−0.15013	−0.12797
SGWJA	0.00526	−0.00217	−0.13575	−0.12264	0.01317	0.00641	−0.15017	−0.12804

Table 1: The bias of the estimates. (LS: least squares estimate, H: Huber's M-estimate, M: Mallows' GM-estimate, S: Schweppe's GM-estimate, OJ: ordinary jackknife estimate, WJ: weighted jackknife estimate, GWJ: general weighted jackknife estimate, A: approxitate procedure)

also can find that it is hard to judge between Mallows' and Schweppe's GM-estimate in comparing the biases of estimates. When we check the biases of the exact jackknife robust estimates and the approximate jackknife robust estimates, we can see that the biases are similar for those two types procedures.

2.4.2 The Relative Biases of Variance-Covariance Estimates

The relative biases of covariance estimates are presented through Table 2 to Table 4. From the simulation results, generally, it is hard to make decision on those estimates. But we can find that the results suggested that the approximate jackknife procedures worked well. Compare the relative bias of exact procedure and approximate procedure, we can see that for the case of $\nu = 0.2$ and $n = 40$, all the corresponding results show that the approximate procedures have small bias. It suggests that when we do robust jackknife estimate, it is not necessary to run delete-one case robust M or GM-estimate, then calculate the corresponding jackknife estimate. We only need apply the approximate formulas to get the jackknife robust estimate. (In comparing two procedures, it is really important for saving computation time.)

As the results show us that it is hard to consider three values of ν_{ij} at same time. Since the slope parameter seems more important in the reality, we put ν_{22} under our consideration. We set the critical values as: for the case of $\nu = 0$, we take $|\gamma_{22}| \leq 10$ and for $\nu = 0.2$, we take $|\gamma_{22}| \leq 20$. From such critical values, we have following summary results.

Type	$n = 20$						$n = 40$					
	$\nu = 0$			$\nu = 0.2$			$\nu = 0$			$\nu = 0.2$		
	γ_{11}	γ_{12}	γ_{22}	γ_{11}	γ_{12}	γ_{22}	γ_{11}	γ_{12}	γ_{22}	γ_{11}	γ_{12}	γ_{22}
H.1.a	−19.4	−16.2	−8.6	−0.5	0.8	1.0	−1.2	0.6	0.5	14.7	18.1	4.7
H.1.b	−18.6	−16.9	−11.9	−32.6	−58.7	−43.3	−1.9	0.4	0.8	−29.7	−64.6	−54.5
H.1.c	−11.4	−9.5	−4.1	−26.6	−55.0	−38.2	2.5	4.9	5.4	−26.5	−63.0	−52.4
H.1.d	−14.1	−11.5	−4.3	−7.4	−17.5	−10.9	1.2	3.1	3.4	0.4	−11.4	−16.1
H.1.e	−17.7	−15.2	−8.3	−11.3	−21.0	−14.6	−1.0	0.8	1.1	−2.0	−13.5	−18.1
H.2.a	2.2	10.1	15.3	−24.2	−54.0	−30.0	9.1	15.7	16.5	−28.6	−69.3	−53.5
H.3.a	6.5	15.9	27.3	32.6	44.8	43.5	14.3	18.0	17.0	33.9	47.7	34.9
H.3.b	−0.9	1.5	11.8	21.3	16.8	15.0	11.9	12.3	11.5	24.9	25.7	13.0
H.3.c	−1.4	3.6	14.0	18.4	16.7	15.3	11.3	13.2	12.3	23.2	25.1	12.6
H.4.a	−3.5	6.8	16.5	37.1	78.2	87.9	6.3	11.4	11.1	43.5	76.6	62.1
H.5.a	−10.5	−6.9	1.6	10.6	12.0	12.2	4.0	5.9	5.8	20.8	24.3	10.2
H.5.b	−9.6	−7.6	−2.1	−25.2	−54.1	−36.9	3.2	5.6	6.1	−26.0	−62.7	−52.1
H.5.c	−1.5	0.6	6.6	−18.5	−50.0	−31.3	7.9	10.5	11.0	−22.6	−61.0	−49.9
H.5.d	−4.6	−1.7	6.4	2.9	−8.3	−1.0	6.6	8.5	8.8	5.7	−6.7	−11.7
H.5.e	−8.6	−5.8	1.9	−1.5	−12.2	−5.2	4.2	6.1	6.4	3.2	−8.9	−13.7
H.6.a	−10.9	−4.9	3.6	8.3	12.4	13.0	3.5	6.7	6.5	19.2	23.9	10.1
H.6.b	−21.5	−19.0	−13.2	−28.7	−47.0	−36.0	−3.9	−2.1	−2.1	−24.7	−52.3	−47.2
H.6.c	2.6	10.7	19.6	8.0	2.5	10.9	9.8	14.6	14.8	7.7	−2.1	−7.1
H.6.d	−1.2	6.7	13.0	−18.6	−39.6	−20.6	6.8	12.4	12.7	−23.3	−56.6	−45.7
H.7.a	−15.1	−13.8	−7.7	−22.3	−45.0	−32.9	0.6	2.1	2.5	−18.9	−47.3	−42.8

Table 2: The relative bias of variance-covariance for Huber's M-estimate case

Type	$n = 20$						$n = 40$					
	$\nu = 0$			$\nu = 0.2$			$\nu = 0$			$\nu = 0.2$		
	γ_{11}	γ_{12}	γ_{22}	γ_{11}	γ_{12}	γ_{22}	γ_{11}	γ_{12}	γ_{22}	γ_{11}	γ_{12}	γ_{22}
M.1.a	−18.9	−15.7	−8.7	0.4	7.8	9.3	−1.3	0.5	0.4	14.1	26.8	16.0
M.1.b	−18.9	−17.1	−12.0	−28.2	−50.7	−32.1	−2.0	0.2	0.5	−21.7	−52.4	−38.0
M.1.c	−11.6	−9.8	−4.1	−21.8	−46.3	−26.1	2.5	4.8	5.1	−18.1	−50.2	−35.1
M.1.d	−13.6	−10.9	−4.0	−3.8	−6.8	1.5	1.2	3.0	3.3	4.5	2.1	0.3
M.1.e	−17.2	−14.6	−8.1	−7.9	−10.8	−2.9	−1.1	0.7	1.0	2.1	−0.2	−2.0
M.2.a	1.0	8.3	13.8	−19.6	−46.5	−17.9	8.8	15.2	15.9	−20.7	−59.2	−37.1
M.3.a	5.4	14.3	25.8	25.3	36.6	38.2	13.8	17.3	16.2	29.8	47.2	34.9
M.3.b	0.3	3.4	13.9	22.9	27.5	28.4	11.9	12.4	11.5	29.3	43.4	31.1
M.3.c	−1.0	4.3	14.8	19.3	26.0	27.4	11.1	12.9	12.0	27.3	42.3	30.3
M.4.a	−4.9	4.4	13.4	13.1	26.8	29.7	5.7	10.4	10.1	20.5	36.9	25.5
M.5.a	−9.9	−6.4	1.5	11.5	19.8	21.4	3.9	5.8	5.6	20.1	33.5	22.1
M.5.b	−9.8	−7.9	−2.2	−20.2	−45.2	−24.6	3.2	5.5	5.8	−17.7	−49.9	−34.7
M.5.c	−1.8	0.3	6.5	−13.1	−40.4	−17.9	7.9	10.3	10.7	−13.8	−47.6	−31.7
M.5.d	−4.0	−0.9	6.7	6.9	3.6	12.8	6.5	8.4	8.7	10.0	7.5	5.6
M.5.e	−8.0	−5.1	2.2	2.4	−0.9	7.9	4.1	6.0	6.3	7.5	5.1	3.2
M.6.a	−11.0	−5.4	2.6	7.8	17.5	19.8	3.1	6.2	6.1	18.2	32.4	19.0
M.6.b	−21.7	−19.4	−13.5	−26.6	−43.2	−29.7	−4.0	−2.3	−2.3	−18.0	−40.5	−32.4
M.6.c	1.8	9.3	17.5	10.3	11.5	21.8	9.3	13.7	14.0	11.4	11.3	9.5
M.6.d	−3.5	3.2	9.6	−17.3	−38.1	−15.9	6.2	11.6	11.3	−16.5	−46.8	−31.3

Table 3: The relative bias of variance-covariance for Mallows' GM-estimate case

Number of cases	Type of estimate	symbol	$\nu = 0$			$\nu = 0.2$		
			γ_{11}	γ_{12} (γ_{21})	γ_{22}	γ_{11}	γ_{12} (γ_{21})	γ_{22}
$n = 20$	least squares	O	−13.537	−14.051	−9.226	−27.912	−59.319	−41.301
		OJ	−5.498	−1.715	5.614	−4.194	−17.787	−20.465
		WJ	−19.829	−23.880	−17.901	−10.677	−29.572	−29.651
		GWJ	−15.299	−15.559	−8.946	−1.892	−13.055	−13.742
	Schweppe's GM	S.1.a	−18.903	−15.471	−7.704	1.484	9.190	10.266
		S.2.a	16.808	25.223	31.044	−7.105	−36.288	−4.425
		S.3.a	8.764	18.603	29.771	27.677	39.447	39.846
		S.3.b	3.275	7.074	17.119	25.391	30.532	30.112
		S.3.c	1.726	7.602	17.666	21.382	28.378	28.511
		S.4.a	−4.104	5.888	15.657	14.818	29.357	31.733
		S.5.a	−9.893	−6.079	2.552	12.759	21.322	22.517
		S.6.a	−10.907	−5.011	3.620	8.947	18.858	20.670
$n = 40$	least squares	O	1.279	3.021	5.862	−27.872	−67.963	−58.057
		OJ	5.824	9.886	13.215	−6.265	−23.828	−32.624
		WJ	−1.904	−2.966	−0.180	−5.053	−28.137	−36.736
		GWJ	0.632	2.009	4.920	5.286	−9.492	−20.212
	Schweppe's GM	S.1.a	−1.567	−0.115	−0.347	15.682	30.011	18.287
		S.2.a	16.971	23.713	24.346	−14.979	−54.562	−31.466
		S.3.a	14.435	18.216	16.901	30.849	48.619	35.765
		S.3.b	12.375	13.058	11.925	30.900	46.039	32.916
		S.3.c	11.619	13.662	12.460	28.642	44.395	31.715
		S.4.a	5.486	9.960	9.551	21.675	39.216	27.101
		S.5.a	3.615	5.142	4.898	21.771	36.853	24.513
		S.6.a	2.883	5.686	5.376	19.635	35.265	23.336

Table 4: The relative bias of variance-covariance for least squares and Schweppe's GM-estimate case

$\nu = 0$ ($\lvert\gamma_{22}\rvert \leq 10$)				$\nu = 0.2$ ($\lvert\gamma_{22}\rvert \leq 20$)			
$n = 20$		$n = 40$		$n = 20$		$n = 40$	
LS	LSOJ	LS	LSWJ	LSGWJ		LSGWJ	
LSGWJ		LSGWJ					
H.1.a	H.1.c	H.1.a	H.1.b	H.1.a	H.1.d	H.1.a	H.1.d
H.1.d	H.1.e	H.1.c	H.1.d	H.1.e	H.3.b	H.1.e	H.3.b
H.5.a	H.5.b	H.1.e	H.5.a	H.3.c	H.5.a	H.3.c	H.5.a
H.5.c	H.5.d	H.5.b	H.5.d	H.5.d	H.5.e	H.5.d	H.5.e
H.5.e	H.6.a	H.5.e	H.6.a	H.6.a	H.6.c	H.6.a	H.6.c
H.7.a		H.6.b	H.7.a				
M.1.a	M.1.c	M.1.a	M.1.b	M.1.a	M.1.d	M.1.a	M.1.d
M.1.d	M.1.e	M.1.c	M.1.d	M.1.e	M.2.a	M.1.e	M.5.d
M.5.a	M.5.b	M.1.e	M.4.a	M.5.c	M.5.d	M.5.e	M.6.a
M.5.c	M.5.d	M.5.a	M.5.b	M.5.e	M.6.d	M.6.c	
M.5.e	M.6.a	M.5.c	M.5.d				
M.6.d		M.5.e	M.6.a				
		M.6.b					
S.1.a	S.5.a	S.1.a	S.4.a	S.1.a	S.2.a	S.1.a	
S.6.a		S.5.a	S.6.a	S.6.a			

Table 5: The recommended procedures based on the results of relative bias of variance of slope parameter

From the summaries of the relative bias of variance estimates, we can see that the values of H.1.a, M.1.a, S.1.a, H.5.a, H.6.a, M.5.a, M.6.a, S.5.a and S.6.a appeare more often than other values. After checking those symbols, we find that the last six values represent the weighted jackknife and general weighted jackknife estimates in the approximate procedure. This supports our proposal. If we check the results carefully, we also can find that the estimators H.2.a, M.2.a and S.2.a seem to have more stable values although they do not have outstanding performance.

2.5 Further Study

In the classical linear regression case, if the normal assumptions hold, then for the estimate $\hat{\beta}$ we have F-statistic

$$F = \frac{(\hat{\beta} - \beta)^T \widehat{cov}(\hat{\beta})(\hat{\beta} - \beta)}{p} , \tag{2.27}$$

where

$$\widehat{cov}(\hat{\beta}) = \hat{\sigma}^2 (X^T X)^{-1} ,$$

which is distributed as F distribution with degree of freedom p and $n-p$ (*i.e.* $F_{p,n-p}$). We can obtain the confidence region of the estimate and make inference about the estimate based on this property.

This nice property is based on the data satisfying the normality assumption. If the data are contaminated, we can not apply this property to make inference. In reality, we will encounter lots of cases when the classical assumptions are not satisfied.

This is the reason why we need to use robust procedures to estimate the parameters. As we know, it is hard to make inferences on the robust estimate even we can use the robust procedure to find the estimators. In our simulation study, we try to figure out a reasonable value of q which will lead to the statistic

$$F = \frac{(\hat{\beta}_r - \beta)^T \widehat{cov(\hat{\beta}_r)} (\hat{\beta}_r - \beta)}{p} , \tag{2.28}$$

having an approximate F distribution with degree of freedoms equal to p and q ($i.e.$ $F_{p,q}$). If we can find the q value for different robust estimate, we can make inference on our robust estimates.

The method which we applied to find the reasonable value of q as following. First we calculate the values of $\hat{F}_1, \hat{F}_2, \ldots, \hat{F}_{no.sim}$ based on above formula, then sorting those value such that $\hat{F}_1^* \leq \hat{F}_2^* \leq \ldots \leq \hat{F}_{no.sim}^*$, those values form the empirical quantiles which can be written as $\hat{F}_{series}^*$. Second, we get the theoretical quantiles values $F_{p,q}^1, F_{p,q}^2, \ldots, F_{p,q}^{no.sim}$ which can be written as $F_{p,q}^{series}$ (where p is the number of parameters which is fixed in our case and $1 \leq q \leq n$) .

We define the correlation between the empirical quantiles and the theoretical quantiles as a function of q which is

$$h(q) = corr(\hat{F}_{series}^*, F_{p,q}^{series}) . \tag{2.29}$$

The estimated value $\hat{q}$ is obtained by condition $h(\hat{q}) = max_{1 \leq q \leq n} h(q)$

Type of	$n = 20$					$n = 40$				
estimate	INIT	OJ	WJ	GWJ	OJA	INIT	OJ	WJ	GWJ	OJA
H.1.a	7 (2)	8 (2)	8 (2)	7 (2)	7 (1)	18 (2)	22 (2)	22 (2)	18 (2)	18 (2)
H.1.b	12 (3)	11 (3)	13 (3)	13 (3)	13 (1)	20 (3)	24 (2)	24 (3)	20 (3)	21 (2)
H.1.c	13 (3)	12 (3)	14 (3)	13 (3)	14 (1)	21 (3)	25 (2)	25 (3)	21 (3)	22 (2)
H.1.d	8 (3)	10 (3)	9 (3)	9 (3)	9 (1)	24 (2)	34 (2)	34 (2)	24 (2)	25 (2)
H.1.e	8 (3)	10 (3)	9 (3)	9 (3)	9 (1)	24 (2)	33 (2)	33 (2)	23 (2)	24 (2)
H.2.a	13 (5)	9 (4)	14 (5)	14 (5)	13 (3)	20 (4)	22 (3)	23 (4)	20 (4)	20 (3)
H.3.a	8 (5)	9 (6)	9 (6)	8 (5)	8 (6)	14 (4)	14 (4)	13 (4)	15 (4)	14 (4)
H.3.b	7 (2)	8 (2)	8 (2)	7 (2)	7 (1)	15 (1)	15 (1)	14 (1)	16 (1)	15 (1)
H.3.c	8 (2)	8 (3)	8 (2)	8 (2)	8 (1)	14 (3)	14 (2)	13 (3)	15 (3)	15 (2)
H.4.a	8 (6)	9 (6)	9 (6)	8 (6)	8 (6)	18 (4)	21 (5)	20 (4)	19 (4)	18 (4)
H.5.a	7 (2)	8 (2)	8 (2)	7 (2)	7 (1)	18 (2)	22 (2)	22 (2)	18 (2)	18 (1)
H.5.b	12 (3)	11 (3)	13 (3)	13 (3)	13 (1)	20 (3)	24 (2)	24 (3)	20 (3)	21 (2)
H.5.c	13 (3)	12 (3)	14 (3)	13 (3)	14 (1)	21 (3)	25 (2)	25 (3)	21 (3)	22 (2)
H.5.d	8 (3)	10 (3)	9 (3)	9 (3)	9 (1)	24 (2)	34 (2)	34 (2)	24 (2)	25 (2)
H.5.e	8 (3)	10 (3)	9 (3)	9 (3)	9 (1)	24 (2)	33 (2)	33 (2)	23 (2)	24 (2)
H.6.a	8 (2)	9 (2)	8 (2)	8 (2)	8 (2)	18 (4)	22 (3)	22 (5)	19 (4)	18 (2)
H.6.b	11 (3)	12 (3)	12 (3)	12 (3)	12 (3)	17 (4)	19 (4)	19 (5)	18 (4)	17 (3)
H.6.c	8 (7)	10 (8)	9 (7)	9 (7)	9 (6)	27 (9)	40 (6)	40 (10)	27 (9)	29 (5)
H.6.d	12 (8)	11 (5)	13 (5)	13 (8)	13 (10)	21 (4)	26 (4)	26 (4)	22 (3)	22 (4)
H.7.a	10 (3)	12 (3)	11 (3)	11 (3)	11 (1)	23 (3)	30 (2)	30 (3)	23 (3)	24 (2)

Table 6: Estimators of $\hat{q}$ for Huber's type M-estimate case, the values in parentheses are the estimators of $\hat{q}$ based on the contamination ratio $\nu = 0.2$, the values outside of the parentheses are the estimators of $\hat{q}$ based on the contamination ratio $\nu = 0$. (INIT: original estimator; OJ: ordinary jackknife estimator; WJ: weighted jackknife estimator; GWJ: general weighted jackknife estimator; OJA: ordinary jackknife estimator using approximate procedure)

Type of estimate	$n = 20$					$n = 40$				
	INIT	OJ	WJ	GWJ	OJA	INIT	OJ	WJ	GWJ	OJA
M.1.a	7 (2)	8 (2)	8 (2)	7 (2)	7 (2)	19 (5)	19 (6)	20 (6)	20 (5)	20 (5)
M.1.b	12 (3)	10 (4)	12 (4)	13 (3)	13 (3)	19 (4)	14 (4)	16 (4)	19 (5)	19 (4)
M.1.c	13 (4)	11 (4)	13 (4)	13 (3)	13 (3)	20 (4)	15 (4)	17 (4)	20 (5)	20 (4)
M.1.d	8 (2)	10 (2)	9 (2)	9 (2)	9 (2)	24 (6)	19 (5)	22 (6)	24 (6)	25 (6)
M.1.e	8 (2)	9 (2)	9 (2)	9 (2)	9 (2)	24 (6)	18 (5)	21 (6)	23 (6)	24 (6)
M.2.a	13 (3)	8 (3)	12 (3)	14 (3)	12 (4)	18 (4)	13 (4)	15 (4)	18 (4)	18 (4)
M.3.a	8 (6)	9 (6)	8 (6)	8 (6)	8 (6)	15 (4)	15 (5)	14 (5)	16 (4)	15 (5)
M.3.b	7 (2)	8 (2)	8 (2)	7 (2)	7 (2)	15 (5)	16 (5)	15 (5)	16 (5)	16 (5)
M.3.c	8 (3)	8 (3)	8 (3)	8 (3)	8 (3)	15 (3)	15 (3)	14 (3)	16 (3)	16 (3)
M.4.a	8 (6)	9 (6)	8 (6)	8 (6)	8 (6)	19 (5)	17 (6)	18 (5)	20 (5)	20 (5)
M.5.a	7 (2)	8 (2)	8 (2)	7 (2)	7 (2)	19 (5)	19 (6)	20 (6)	20 (5)	20 (5)
M.5.b	12 (3)	10 (4)	12 (4)	13 (3)	13 (3)	19 (4)	14 (4)	16 (4)	19 (5)	19 (4)
M.5.c	13 ((4)	11 (4)	13 (4)	13 (3)	13 (3)	20 (4)	15 (4)	17 (4)	20 (5)	20 (4)
M.5.d	8 (2)	10 (2)	9 (2)	9 (2)	9 (2)	24 (6)	19 (5)	22 (6)	24 (6)	25 (6)
M.5.e	8 (2)	9 (2)	9 (2)	9 (2)	9 (2)	24 (6)	18 (5)	21 (6)	23 (6)	24 (6)
M.6.a	8 (4)	8 (3)	8 (4)	8 (3)	8 (3)	19 (5)	18 (6)	19 (6)	20 (5)	20 (5)
M.6.b	11 (4)	11 (4)	11 (4)	12 (5)	11 (5)	17 (5)	14 (4)	15 (5)	18 (5)	18 (5)
M.6.c	9 (4)	10 (5)	9 (5)	9 (4)	9 (4)	27 (8)	19 (8)	23 (9)	27 (8)	28 (8)
M.6.d	12 (4)	10 (4)	12 (3)	13 (4)	13 (4)	20 (5)	15 (4)	17 (5)	21 (5)	21 (4)

Table 7: Estimators of $\hat{q}$ for Mallows' type GM-estimate case, the values in parentheses are the estimators of $\hat{q}$ based on the contamination ratio $\nu = 0.2$, the values outside of parentheses are the estimators of $\hat{q}$ based on the contamination ratio $\nu = 0$. (INIT: original estimator; OJ: ordinary jackknife estimator; WJ: weighted jackknife estimator; GWJ: general weighted jackknife estimator; OJA: ordinary jackknife estimator using approximate procedure)

Number of cases	Type of estimate	$\nu = 0$					$\nu = 0.2$				
		INIT	OJ	WJ	GWJ	OJA	INIT	OJ	WJ	GWJ	OJA
$n = 20$	S.1.a	7	8	8	7	7	1	1	1	1	1
	S.2.a	12	4	6	12	13	2	2	2	2	2
	S.3.a	8	9	8	8	8	5	6	6	5	5
	S.3.b	7	8	7	7	7	1	1	1	1	1
	S.3.c	8	8	8	8	8	1	1	1	1	1
	S.4.a	8	8	8	8	8	6	6	6	6	6
	S.5.a	7	8	8	7	7	1	1	1	1	1
	S.6.a	8	8	8	7	8	2	2	2	2	2
$n = 40$	S.1.a	17	18	19	17	17	4	4	4	4	4
	S.2.a	15	13	16	15	14	3	3	3	3	3
	S.3.a	13	14	13	13	13	3	4	3	3	4
	S.3.b	13	15	14	14	14	4	4	4	4	4
	S.3.c	13	15	14	14	14	4	4	4	4	4
	S.4.a	16	18	19	17	17	4	4	4	4	4
	S.5.a	17	18	19	17	17	4	4	4	4	4
	S.6.a	17	18	19	17	17	4	4	4	4	4

Table 8: Estimators of $\hat{q}$ for Schweppe's type GM-estimate

The simulation results of $\hat{q}$ are shown on Table 6 to Table 8. From those results, we can see that when the data are contaminated, the $\hat{q}$ values tend to be small. The majority values are between 3 to 6 for the case of $\nu = 0.2$. For the case of $\nu = 0$, the $\hat{q}$ values are larger than those obtained from $\nu = 0.2$. We also can see that the variation of $\hat{q}$ values is large in the case of $\nu = 0$.

3 Diagnostic Statistics for Robust Estimate

3.1 A Brief Review of Diagnostic Statistics for Ordinary Least Squares Estimates

3.1.1 Leverage and Residuals

In diagnostic procedures, the hat matrix H plays an important role. The diagonal elements of the hat matrix also are called leverages. We mentioned before that the hat matrix has several properties, i.e. it is an idempotent matrix and the diagonal elements of H have $\sum_{i=1}^{n} h_i = p$, the average size of a diagonal element is p/n , we also can see that the leverages have $0 \leq h_i \leq 1$. (Actually, the h_i satisfy $1/n \leq h_i \leq 1$, See Belsley et al (1980)).

C. R. Rao (1973) gave the following result. If the rows of $\tilde{X}$ (which is the centred matrix of X) are assumed to be *i.i.d.* from a $(p-1)$ dimensional normal distribution, then

$$\frac{n-p}{p-1}[\frac{1-\Lambda(\tilde{\mathbf{x}}_i)}{\Lambda(\tilde{\mathbf{x}}_i)}] \sim F_{p-1,n-p},$$

where $\tilde{\mathbf{x}}_i$ are centred vector of $\mathbf{x}_i$ and $\Lambda(\tilde{\mathbf{x}}_i)$ is the Wilks' statistic which has the expression

$$\Lambda(\tilde{\mathbf{x}}_i) = \frac{|\tilde{X}^T\tilde{X} - (n-1)\bar{\tilde{\mathbf{x}}}_{(i)}\bar{\tilde{\mathbf{x}}}_{(i)}^T - \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T|}{|\tilde{X}^T\tilde{X}|} .$$

where $\bar{\tilde{\mathbf{x}}}_{(i)}$ is the mean value of centered vector $\tilde{\mathbf{x}}_j$ when $\tilde{\mathbf{x}}_i$ is deleted, *i.e.*

$$\bar{\tilde{\mathbf{x}}}_{(i)} = \frac{1}{n-1}\sum_{j\neq i}^{n}\tilde{\mathbf{x}}_j$$

It can be simplified as :

$$\Lambda(\tilde{\mathbf{x}}_i) = \frac{n}{n-1}(1-h_i)\ , \tag{3.1}$$

so, we have that

$$\frac{(n-p)(h_i-\frac{1}{n})}{(1-h_i)(p-1)} \sim F_{p-1,n-p}\ .$$

This result is useful for us to choose cutoff value to detect the leverage point. For $p > 10$ and $n - p > 50$ the 95% value for F is less than 2 and hence $2p/n$ (twice the balanced average h_i) is a good rough cutoff. When $p/n > 0.4$, there are so few degree of freedom per parameter that all observations become suspect. For small p , $2p/n$ tends to call a few too many points to our attention. Since it is simple to use and easy to remember, it is suggested that $2p/n$ as a cutoff. When h_i exceeds $2p/n$, we call the ith observation a *leverage point.*

Same as the leverage h_i , the residuals r_i also play an important role in statistical diagnostic. For the residuals, there are two kinds of modification. One is *standardized residuals*, which can be expressed as

$$r_{si} = \frac{r_i}{\hat{\sigma}\sqrt{1-h_i}}\ , \tag{3.2}$$

where $\hat{\sigma}^2 = MSE$. The other one is *studentized residuals*, which have the expression

$$r_i^* = \frac{r_i}{\hat{\sigma}_{(i)}\sqrt{1-h_i}}\ , \tag{3.3}$$

where $\hat{\sigma}^2_{(i)} = MSE_{(i)}$. It is the error mean square calculated when the ith observation is omitted.

The studentized residual, in a number of practical situations, is distributed closely to the t-distribution with $n - p - 1$ degree of freedom. From this point of view, we can readily assess the significance of any single studentized residual. The studentized residuals thus provide a better way to examine the information in the residuals. But it is important to note that some influential points can have relatively small studentized residuals. Indeed, this point is central to the notion of *Bounded Influence M-estimation*, i.e. Generalized M-estimate or write in short as GM-estimate, in which the more influential points are automatically down-weighted.

From the leverage and the scaled residuals, we can form several diagnostic statistics. It is easy to see that the diagnostic statistics we will mention later are basically formed by different combinations of functions of leverages h_i and scaled residuals although they are obtained under the approach of the influence function and a class of norms which are location and scale invariant.

3.1.2 The Empirical Influence Curve and the Sample Influence Curve

To measure the influence of each observation in a data set, we can also use the idea of the influence function. The influence function of a statistic $T(F)$ at a point z_i, $IF(z_i; T, F)$, measures the influence on the statistic of adding one observation $z_i = (\mathbf{x}_i, y_i)$ to a large sample. For ordinary least squares regression, the empirical influence curve and the sample influence curve can be obtained in the following way (See Chatterjee and Hadi (1986)).

(1). Let $\hat{F}_n$ be the empirical distribution function based on the full sample and $\hat{F}_{(i)}$ be the empirical distribution function when the i-th observation is omitted. When we replace F by $\hat{F}_{(i)}$ and $T(\hat{F}_{(i)})$ by $\hat{\beta}_{(i)}$, we can obtain the empirical influence curve which is

$$EIC_i = (n-1)(X_{(i)}^T X_{(i)})^{-1}\mathbf{x}_i(y_i - \mathbf{x}_i^T\hat{\beta}_{(i)}) \ . \tag{3.4}$$

Since in the ordinary least squares case, we have following formula

$$(X_{(i)}^T X_{(i)})^{-1}\mathbf{x}_i = \frac{(X^TX)^{-1}\mathbf{x}_i}{1-h_i} \ ,$$

and

$$r_{i(i)} = y_i - \mathbf{x}_i^T\hat{\beta}_{(i)} = \frac{y_i - \mathbf{x}_i^T\hat{\beta}}{1-h_i} = \frac{r_i}{1-h_i} \ , \tag{3.5}$$

the value of EIC_i in the ordinary least squares case can be expressed as

$$EIC_i = (n-1)(X^TX)^{-1}\mathbf{x}_i\frac{r_i}{(1-h_i)^2} \ . \tag{3.6}$$

(2). When we omit the limit in the expression of the influence function and take $F = \hat{F}_n$, $T(\hat{F}_n) = \hat{\beta}$ and $\varepsilon = -1/(n-1)$, we obtain the sample influence curve (SIC) which is:

$$\begin{aligned} SIC_i &= (n-1)(X^TX)^{-1}\mathbf{x}_i(y_i - \mathbf{x}_i^T\hat{\beta}_{(i)}) \\ &= (n-1)(X^TX)^{-1}\mathbf{x}_i\frac{r_i}{(1-h_i)} \\ &= (n-1)(\hat{\beta} - \hat{\beta}_{(i)}) \end{aligned}$$

The last equation is obtained by the formula (1.22).

From above formulas of EIC_i and SIC_i, we can see that EIC_i is more sensitive to h_i while SIC_i is proportional to the distance between $\hat{\beta}$ and $\hat{\beta}_{(i)}$.

Since the influence function $IF(\mathbf{x}_i, y_i; F, T)$ is a vector, it has to be normalized so that observations can be ordered in a meaningful way (See Chatterjee and Hadi (1986)). The class of norms which are location and scale invariant is given by

$$D_i(A; c) = \frac{(IF_i)^T A (IF_i)}{c} \ . \tag{3.7}$$

For any appropriate choice of matrix A and scalar c, a large value of $D_i(A; c)$ indicates that the ith observation has strong influence on the estimated coefficients relative to A and c. Following are some commonly suggested choices of A, c and their corresponding diagnostic statistics.

3.1.3 Cook's Distance

If we use the sample influence curve to approximate the influence function and substitute $A = X^T X$ and $c = (n-1)^2 p\hat{\sigma}^2$, we can get

$$\begin{aligned} C_i &= D_i(X^T X; (n-1)^2 p\hat{\sigma}^2) = \frac{(\hat{\beta} - \hat{\beta}_{(i)})^T X^T X (\hat{\beta} - \hat{\beta}_{(i)})}{p \cdot MSE} \\ &= \frac{(\hat{y} - \hat{y}_{(i)})^T (\hat{y} - \hat{y}_{(i)})}{p \cdot MSE} \ . \end{aligned}$$

It can be written in short form

$$C_i = \frac{r_{si}^2}{p} \frac{h_i}{(1 - h_i)} \ . \tag{3.8}$$

Note that $\hat{y}_{(i)}$ is the vector of predicted values when $y_{(i)}$ is regressed on $X_{(i)}$. Thus C_i can be interpreted as the scaled Euclidean distance between the two vectors of fitted values when the fitting is done by including or excluding the ith observation. This measure was proposed by Cook (1977). It is called *Cook's distance*. Cook suggested that each C_i be compared with the quantiles of the central F distribution with p and $(n-p)$ degree of freedom. If the percentile value is near 50 percent or more, the distance between the vector $\hat{\beta}$ and $\hat{\beta}_{(i)}$ should be considered large, implying that the i-th case has a substantial influence on the fit of the regression function.

3.1.4 Welsch-Kuh Distance ($DFFITS$)

We can use $|\hat{y}_i - \hat{y}_{(i)}|/\sqrt{MSE \cdot h_i}$ to measure the impact of the ith observation on the ith predicted value. Welsch and Kuh (1977), Welsch and Peters (1978), and Belsley, Kuh and Welsch (1980) suggested using $\hat{\sigma}^2_{(i)}$ as an estimate of σ^2. This statistic is called as *Welsch-Kuh distance*. It is also called $DFFITS$ ($DFFITS$ stands for scaled difference of fits) which can be expressed as

$$WK_i = DFFITS_i = \frac{|\hat{y}_i - \hat{y}_{(i)}|}{\sqrt{MSE_{(i)} \cdot h_i}} = \frac{|\mathbf{x}_i(\hat{\beta} - \hat{\beta}_{(i)})|}{\sqrt{MSE_{(i)} \cdot h_i}} ,$$

We have following simple expression for $MSE_{(i)}$:

$$MSE_{(i)} = \frac{(n-p) \cdot MSE}{n-p-1} - \frac{r_i^2}{(n-p-1) \cdot (1-h_i)} .$$

After squaring, $DFFITS_i$ is equivalent to the value of $D_i(A;\ c)$ when the sample influence curve is used to approximate the influence function and take $A = X^T X$,

$c = (n-1)^2 \cdot MSE_{(i)}$, *i.e.*

$$WK_i = DFFITS_i = \sqrt{D_i(X^TX;\ (n-1)^2 \cdot MSE_{(i)})} = |r_i^*|\sqrt{\frac{h_i}{1-h_i}}\ . \qquad (3.9)$$

For a perfectly balanced design matrix X ($h_i = p/n$ for all i), we have

$$WK_i = DFFITS_i = (\frac{p}{n-p})^{1/2} r_i^*\ .$$

Belsley, Kuh and Welsch (1980) suggested using $2 \cdot \sqrt{p/n}$ as a calibration point for $DFFITS_i$. Velleman and Welsch (1981) recommended that if the $DFFITS_i$ value is greater than 1 or 2, then the correspondence observations are worthy of special attention.

3.1.5 Welsch's Distance

When we use the empirical influence curve (EIC) to approximate the influence function and set $A = X_{(i)}^T X_{(i)}$, $c = (n-1) \cdot MSE_{(i)}$, we can get

$$W_i^2 = (n-1) r_i^{*2} \frac{h_i}{(1-h_i)^2}\ . \qquad (3.10)$$

Comparing (3.9) with (3.10), we have that

$$W_i = DFFITS_i \sqrt{\frac{n-1}{1-h_i}} = WK_i \sqrt{\frac{n-1}{1-h_i}}\ .$$

W_i is called *Welsch's distance*. This measure was suggested by Welsch (1982). It also can be used as a diagnostic tool. For the sample size $n > 15$, we can use $3\sqrt{p}$ as a calibration point for W_i.

3.2 The Diagnostic Statistics of Robust Estimate

Analogous to the case of ordinary least squares, we propose several measures which can be used to detect possible outliers and influential points in robust regression problems. As in section one, we consider the robust estimators $\hat{\beta}_\phi = T(\hat{F}_n)$ and $\hat{\beta}_{\phi(i)} = T(\hat{F}_{(i)})$ which satisfy the general functional forms $\sum_{j=1}^{n} \phi(z_j; \hat{\beta}_\phi, \hat{\sigma}_\phi) = \mathbf{0}$ and $\sum_{j\neq i}^{n} \phi(z_j; \hat{\beta}_{\phi(i)}, \hat{\sigma}_{\phi(i)}) = \mathbf{0}$.

3.2.1 Robust Leverage and Robust Mahalanobis Distance

We name the w_i^ϕ which has the expression in formula (1.44) *robust leverages* in analogy to the leverages in the case of ordinary least squares. For the robust leverages, if ordinary least squares is applied, we have $w_i^\phi = h_i$ (which we mentioned before). Robust leverages can be used as one of the diagnostics.

When measuring the distance between two observations $\mathbf{x}$ and $\mathbf{y}$ in a space, we can use *Mahalanobis distance*. By the definition, the square root of

$$\Delta_\Sigma^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \Sigma^{-1} (\mathbf{x} - \mathbf{y}) \,, \tag{3.11}$$

is called the Mahalanobis distance between $\mathbf{x}$ and $\mathbf{y}$, under matrix Σ .

Similar to the Mahalanobis distance defined above, we define the square root of $\mathbf{x}_i^T \widehat{M}_\phi^{-1} \mathbf{x}_i$ as the *robust Mahalanobis distance*. Using the robust Mahalanobis distance, we can check the distances between observation $\mathbf{x}_i$ ($i = 1, \ldots, n$) and the rest of observations in a more robust way.

3.2.2 Robustly Standardized Residuals and Deleted Residuals

For the robust estimate case, we define *robustly standardized residuals* r_{si}^{ϕ} by

$$r_{si}^{\phi} = \frac{r_i^{\phi}}{\hat{\sigma}_{\phi} \cdot \sqrt{1 - w_i^{\phi}}} \,. \tag{3.12}$$

In the ordinary least squares case, the robustly standardized residuals r_{si}^{ϕ} becomes the standardized residuals r_{si} as we take $w_i^{\phi} = h_i$ for the ordinary least squares case. Here we need to mention that for ordinary M-estimates, this statistic may not provide sufficient evidence to detect the unusual point of design matrix since the M-estimate is not robust against outliers in the regressors (*i.e.* the X).

We also can use the robust delete-one residuals $r_{(i)}^{\phi}$ as our diagnostic statistics. We define *robust deleted residuals* as

$$r_{d(i)}^{\phi} = \frac{r_{(i)}^{\phi}}{\hat{\sigma}_{\phi}} = \frac{\mathbf{y}_i - \mathbf{x}_i^T \hat{\beta}_{\phi(i)}}{\hat{\sigma}_{\phi}} \,. \tag{3.13}$$

Using this statistic, we can check the influence of each observation on the residuals.

When $\hat{\beta}_{\phi(i)}$ is approximated by $\hat{\beta}_{\phi(i)}^{a}$, we have the expressions of deleted residuals. For the M-estimate, the deleted residuals are

$$r_{d(i)}^{M} \approx \frac{\mathbf{y}_i - \mathbf{x}_i^T \hat{\beta}_{M(i)}}{\hat{\sigma}_M} = \frac{\mathbf{y}_i - \mathbf{x}_i^T \hat{\beta}_{M(i)}^{a}}{\hat{\sigma}_M}$$

which can be simplified as

$$r_{d(i)}^{M} \approx \frac{r_i^M}{\hat{\sigma}_M}[1 + \frac{wt_i^M \mathbf{x}_i^T \widehat{M}_M^{-1} \mathbf{x}_i}{1 - w_i^M}] \,, \tag{3.14}$$

where

$$wt_i^M = \frac{\psi(r_i^M/\hat{\sigma}_M)}{r_i^M/\hat{\sigma}_M} .$$

For the GM-estimate, we have the expression for $r_{d(i)}^{GM}$ which are

$$r_{d(i)}^{GM} \approx \frac{r_i^{GM}}{\hat{\sigma}_{GM}}[1 + \frac{wt_i^{GM}\mathbf{x}_i^T \widehat{M}_{GM}^{-1}\mathbf{x}_i}{1 - w_i^{GM}}] , \tag{3.15}$$

where

$$wt_i^{GM} = \frac{\eta(\mathbf{x}_i, r_i^{GM}/\hat{\sigma}_{GM})}{r_i^{GM}/\hat{\sigma}_{GM}} .$$

Note: If the ordinary least squares estimate is used, the deleted residuals are the exact values of deleted residuals, *i.e.*

$$r_{d(i)}^{LS} = r_i[1 + \frac{h_i}{1 - h_i}] = \frac{r_i}{1 - h_i} ,$$

The above result is obtained since in the ordinary least squares case $wt_i^{\phi} = 1$, $\widehat{M}_{\phi} = X^T X$ and $w_i^{\phi} = 1$.

3.2.3 The Standardized Change in Fit

We define the standardized change in fit which measures the difference of the fits, i.e. the difference between the fit when the i-th observation is present and the fit when the i-th observation is deleted. We use SCF_i to represent it. SCF_i can be written as

$$SCF_i^{\phi} = |\frac{\hat{y}_i^{\phi} - \hat{y}_{(i)}^{\phi}}{\hat{\sigma}_{\phi}}| = |\frac{\mathbf{x}_i^T(\hat{\beta}_{\phi} - \hat{\beta}_{\phi(i)})}{\hat{\sigma}_{\phi}}| .$$

If $\hat{\beta}_{\phi(i)}$ is replaced by $\hat{\beta}^{a}_{\phi(i)}$, we can get the SCF_i values in the simple form. For the M-estimate case, the expression is

$$SCF_i^M = wt_i^M \cdot |r_{si}^M| \cdot \frac{\mathbf{x}_i^T \widehat{M}_M^{-1} \mathbf{x}_i}{\sqrt{1 - w_i^M}} \ . \tag{3.16}$$

Similarly, we have the standardized change in fit for GM-estimate which is

$$SCF_i^{GM} = wt_i^{GM} \cdot |r_{si}^{GM}| \cdot \frac{\mathbf{x}_i^T \widehat{M}_{GM}^{-1} \mathbf{x}_i}{\sqrt{1 - w_i^{GM}}} \ . \tag{3.17}$$

Note: Compare the formula of standardized change of fit (SCF_i) to the Welsch-Kuh distance, we can see that if we divided SCF_i by the square root of robust leverage w_i^{ϕ}, the statistic we obtained is an analogue to the Welsch-Kuh distance. Since some of w_i^{ϕ} might be equal to zero, we prefer to choose SCF_i instead to define a statistic which is an analogue of Welsch-Kuh distance.

3.2.4 The Empirical Influence Curve in the Robust Estimate Case

From the formula for the influence function, we can get the empirical influence curve (EIC_i^{ϕ}) which is found by substituting $\hat{F}_{(i)}$ for F and $\hat{\beta}_{\phi(i)}$ for $T(\hat{F}_{(i)})$. The general form of EIC_i^{ϕ} can be expressed as

$$EIC_i^{\phi} = -(\frac{1}{n-1} \sum_{j \neq i} \frac{\partial}{\partial \hat{\beta}_{\phi(i)}} \phi(z_j; \hat{\beta}_{\phi(i)}, \hat{\sigma}_{\phi(i)}))^{-1} \phi(z_i, \hat{\beta}_{\phi(i)}, \hat{\sigma}_{\phi(i)}). \tag{3.18}$$

As explained earlier, we replace $\hat{\sigma}_{\phi(i)}$ by $\hat{\sigma}_{\phi}$ and the above formula becomes

$$EIC_i^{\phi} = -(n-1)(\sum_{j \neq i} \frac{\partial}{\partial \hat{\beta}_{\phi(i)}} \phi(z_j; \hat{\beta}_{\phi(i)}, \hat{\sigma}_{\phi}))^{-1} \phi(z_i, \hat{\beta}_{\phi(i)}, \hat{\sigma}_{\phi}) \ .$$

In the case of M-estimate, the empirical influence curve EIC_i^M is

$$EIC_i^M = (n-1)(\sum_{j\neq i} \mathbf{x}_j \mathbf{x}_j^T \psi'(\frac{y_j - \mathbf{x}_j^T \hat{\beta}_{M(i)}}{\hat{\sigma}_M})\frac{1}{\hat{\sigma}_M})^{-1} \psi(\frac{y_j - \mathbf{x}_j^T \hat{\beta}_{M(i)}}{\hat{\sigma}_M})\mathbf{x}_i \ .$$

Write in short, the empirical influence curve EIC_i^M can be written as

$$EIC_i^M = (n-1)\hat{\sigma}_M \widehat{M}_{M(i)}^{-1} \psi(\frac{r_{i(i)}^M}{\hat{\sigma}_M})\mathbf{x}_i \ , \tag{3.19}$$

where

$$\widehat{M}_{M(i)} = \sum_{j\neq i} \psi'(\frac{r_{j(i)}^M}{\hat{\sigma}_M})\mathbf{x}_j \mathbf{x}_j^T \ , \tag{3.20}$$

Similar to the M-estimate case, we can get the empirical influence curve for GM-estimate (EIC_i^{GM}) which can be written as

$$EIC_i^{GM} = (n-1)\hat{\sigma}_{GM} \widehat{M}_{GM(i)}^{-1} \eta(\mathbf{x}_j, \frac{r_{i(i)}^{GM}}{\hat{\sigma}_{GM}})\mathbf{x}_i \ , \tag{3.21}$$

where

$$\widehat{M}_{GM(i)} = \sum_{j\neq i} \eta'(\mathbf{x}_j, \frac{r_{j(i)}^{GM}}{\hat{\sigma}_{GM}})\mathbf{x}_j \mathbf{x}_j^T \ . \tag{3.22}$$

When we calculate the EIC_i^M or EIC_i^{GM} values, we can use $\hat{\beta}_{M(i)}^a$ and $\hat{\beta}_{GM(i)}^a$ to approximate $\hat{\beta}_{M(i)}$ and $\hat{\beta}_{GM(i)}$. It will also save a lot of computing time.

3.2.5 The Sample Influence Curve in the Robust Estimate Case

If we omit the limit in the expression for the influence function and take $F = \hat{F}_n$, $T(\hat{F}_n) = \hat{\beta}_\phi$, and $\varepsilon = -1/(n-1)$, we can get the sample influence curve (SIC_i^ϕ) which is

$$SIC_i^\phi = (n-1)(T(\hat{F}_n) - T(\hat{F}_{(i)}))$$

$$= (n-1)(\hat{\beta}_{\phi} - \hat{\beta}_{\phi(i)})$$

$$= (n-1)\delta_{\phi(i)} .$$

Compare it with (1.68), SIC_i^{ϕ} can be written as

$$SIC_i^{\phi} \approx -(n-1)(\sum_{j \neq i}^{n} \phi'(z_j; \hat{\beta}_{\phi}, \hat{\sigma}_{\phi}))^{-1} \phi(z_i; \hat{\beta}_{\phi}, \hat{\sigma}_{\phi}) . \tag{3.23}$$

In the case of M-estimate, SIC_i^M becomes

$$SIC_i^M \approx (n-1)\hat{\sigma}_M \frac{\psi(r_i^M/\hat{\sigma}_M)}{1 - w_i^M} \widehat{M}_M^{-1} \mathbf{x}_i . \tag{3.24}$$

In the GM-estimate case, SIC_i^{GM} can be obtained by

$$SIC_i^{GM} \approx (n-1)\hat{\sigma}_{GM} \frac{\eta(\mathbf{x}_i, r_i^{GM}/\hat{\sigma}_{GM})}{1 - w_i^{GM}} \widehat{M}_{GM}^{-1} \mathbf{x}_i . \tag{3.25}$$

3.2.6 The Diagnostic Statistics Based on The Influence Function

For robust M and GM-estimate, we already have the expressions for $\widehat{M}_M$, $\widehat{M}_{GM}$, $\widehat{M}_{M(i)}$, $\widehat{M}_{GM(i)}$, $\hat{Q}_M$ and $\hat{Q}_{GM}$ before (see (1.50), (1.55), (3.20), (3.22), (1.90), and (1.93)). Now we give the exprssions for $\hat{Q}_{M(i)}$ and $\hat{Q}_{GM(i)}$.

$$\hat{Q}_{M(i)} = \sum_{j \neq i} \psi^2(\frac{r_{j(i)}^M}{\hat{\sigma}_M}) \mathbf{x}_j \mathbf{x}_j^T , \tag{3.26}$$

$$\hat{Q}_{GM(i)} = \sum_{j \neq i} \eta^2(\frac{\mathbf{x}_j, r_{j(i)}^{GM}}{\hat{\sigma}_{GM}}) \mathbf{x}_j \mathbf{x}_j^T . \tag{3.27}$$

Considering the class of norms

$$D_i^{\phi}(A; c) = \frac{(IF_i^{\phi})^T A (IF_i^{\phi})}{c}, \tag{3.28}$$

when we use either the empirical influence curve (EIC_i^M, EIC_i^{GM}) or sample influence curve (SIC_i^M, SIC_i^{GM}) to approximate the influence function IF_i^ϕ and use one of $\widehat{M}_\phi$, $\widehat{M}_{\phi(i)}$, $\widehat{Q}_\phi$, or $\widehat{Q}_{\phi(i)}$ to replace matrix A and choose the appropriate scale value c, we can form some diagnostic statistics for robust estimate. We can see that those statistics are the analogue to the diagnostic statistics in the ordinary least squares case.

3.2.6.1 Robust Analogue of Cook's Distance

If we choose sample influence curve (SIC_i^ϕ) to approximate the influence function IF_i^ϕ and take $c = (n-1)^2 p\hat{\sigma}_\phi^2$ and use $\widehat{M}_\phi$ or $\widehat{M}_\phi \widehat{Q}_\phi^{-1} \widehat{M}_\phi$ to replace matrix A respectively, we have following statistics.

(1). When $A = \widehat{M}_\phi$ and $c = (n-1)^2 p\hat{\sigma}_\phi^2$, the diagnostic statistic becomes

$$D_i^\phi(\widehat{M}_\phi;\ (n-1)^2 p\hat{\sigma}_\phi^2) = \frac{(SIC_i^\phi)^T \cdot \widehat{M}_\phi \cdot SIC_i^\phi}{(n-1)^2 p\hat{\sigma}_\phi^2} \ . \tag{3.29}$$

We name this diagnostic statistic *Cook's distances for robust estimate* and use C_i^ϕ to represent it. In the case of M-estimate, the diagnostic statistic is

$$C_i^M = D_i^M(\widehat{M}_M;\ (n-1)^2 p\hat{\sigma}_M^2) = \frac{(SIC_i^M)^T \cdot \widehat{M}_M \cdot SIC_i^M}{(n-1)^2 p\hat{\sigma}_M^2} \ .$$

Using formulas (1.50) and (3.24), the above expression can be written as

$$C_i^M \approx \frac{\psi^2(r_i^M/\hat{\sigma}_M)}{p} \frac{\mathbf{x}_i^T \widehat{M}_M^{-1} \mathbf{x}_i}{(1-w_i^M)^2} \tag{3.30}$$

Similarly, we can get the statistic for the case of GM-estimate, which is

$$C_i^{GM} \approx \frac{\eta^2(\mathbf{x}_i, r_i^{GM}/\hat{\sigma}_{GM})}{p} \frac{\mathbf{x}_i^T \widehat{M}_{GM}^{-1} \mathbf{x}_i}{(1 - w_i^{GM})^2} \tag{3.31}$$

Note: We can see that there is a relationship between SCF_i^{ϕ} and C_i^{ϕ}.

$$SCF_i^{\phi} = \sqrt{p \cdot C_i^{\phi} \cdot \mathbf{x}_i^T \widehat{M}_{\phi}^{-1} \mathbf{x}_i} \, .$$

(2). $A = \widehat{M}_{\phi} \hat{Q}_{\phi}^{-1} \widehat{M}_{\phi}$ and $c = (n-1)^2 p \hat{\sigma}_{\phi}^2$, the diagnostic statistic becomes

$$D_i^{\phi}(\widehat{M}_{\phi} \hat{Q}_{\phi}^{-1} \widehat{M}_{\phi};\ (n-1)^2 p \hat{\sigma}_{\phi}^2) = \frac{(SIC_i^{\phi})^T \cdot \widehat{M}_{\phi} \hat{Q}_{\phi}^{-1} \widehat{M}_{\phi} \cdot SIC_i^{\phi}}{(n-1)^2 p \hat{\sigma}_{\phi}^2} \, .$$

We name this diagnostic statistic *modified Cook's distances for robust estimate* and write as $\tilde{C}_i^{\phi}$. In the M-estimate case, above formula becomes

$$\tilde{C}_i^M = D_i^M(\widehat{M}_M \hat{Q}_M^{-1} \widehat{M}_M;\ (n-1)^2 p \hat{\sigma}_M^2) = \frac{(SIC_i^M)^T \cdot \widehat{M}_M \hat{Q}_M^{-1} \widehat{M}_M \cdot SIC_i^M}{(n-1)^2 p \hat{\sigma}_M^2} \, .$$

Compare it with formulas (1.50), (3.24), it can be simplified as

$$\tilde{C}_i^M \approx \frac{\psi^2(r_i^M/\hat{\sigma}_M)}{p} \frac{\mathbf{x}_i^T \hat{Q}_M^{-1} \mathbf{x}_i}{(1 - w_i^M)^2} \tag{3.32}$$

For the GM-estimate case, the corresponding diagnostic statistic is

$$\tilde{C}_i^{GM} \approx \frac{\eta^2(\mathbf{x}_i, r_i^{GM}/\hat{\sigma}_{GM})}{p} \frac{\mathbf{x}_i^T \hat{Q}_{GM}^{-1} \mathbf{x}_i}{(1 - w_i^{GM})^2} \, . \tag{3.33}$$

3.2.6.2 Robust Analogue to the Welsch's Distance

If we use empirical influence curve (EIC_i^{ϕ}) to approximate the influence function and choose $A = \widehat{M}_{\phi(i)}$ or $A = \widehat{M}_{\phi(i)}\widehat{Q}_{\phi(i)}^{-1}\widehat{M}_{\phi(i)}$ and take $c = (n-1)\hat{\sigma}_{\phi}^2$, then the diagnostic statistics have the following expressions.

(1). $A = \widehat{M}_{\phi(i)}$ and $c = (n-1)\hat{\sigma}_{\phi}^2$, the diagnostic statistic becomes

$$D_i^{\phi}(\widehat{M}_{\phi(i)};\ (n-1)\hat{\sigma}_{\phi}^2) = \frac{(EIC_i^{\phi})^T \widehat{M}_{\phi(i)} EIC_i^{\phi}}{(n-1)\hat{\sigma}_{\phi}^2} .$$

We name this diagnostic statistic *Welsch's distance for robust estimate* and write it as $W_i^{\phi 2}$. In the case of M-estimate, the statistic is

$$W_i^{M2} = D_i^{M}(\widehat{M}_{M(i)};\ (n-1)\hat{\sigma}_{M}^2) = \frac{(EIC_i^{M})^T \widehat{M}_{M(i)} EIC_i^{M}}{(n-1)\hat{\sigma}_{M}^2} .$$

Comparing it with formulas (3.19 and (3.20)), the above expression can be simplified as

$$W_i^{M2} = (n-1)\psi^2(\frac{r_{i(i)}^{M}}{\hat{\sigma}_M})\mathbf{x}_i^T \widehat{M}_{M(i)}^{-1}\mathbf{x}_i . \tag{3.34}$$

In the case of GM-estimate, it has the expression

$$W_i^{GM2} = (n-1)\eta^2(\mathbf{x}_i, \frac{r_{i(i)}^{GM}}{\hat{\sigma}_{GM}})\mathbf{x}_i^T \widehat{M}_{GM(i)}^{-1}\mathbf{x}_i .$$

(2). $A = \widehat{M}_{\phi(i)}\widehat{Q}_{\phi(i)}^{-1}\widehat{M}_{\phi(i)}$ and $c = (n-1)\hat{\sigma}_{\phi}^2$, the diagnostic statistic is

$$D_i^{\phi}(\widehat{M}_{\phi(i)}\widehat{Q}_{\phi(i)}^{-1}\widehat{M}_{\phi(i)};\ (n-1)\hat{\sigma}_{\phi}^2) = \frac{(EIC_i^{\phi})^T \cdot \widehat{M}_{\phi(i)}\widehat{Q}_{\phi(i)}^{-1}\widehat{M}_{\phi(i)} \cdot EIC_i^{\phi}}{(n-1)\hat{\sigma}_{\phi}^2} .$$

We name the diagnostic statistic obtained based on above formula *Modified Welsch's distance for robust estimate* and express it as $\widetilde{W}_i^{\phi}$. In M-estimate case, the diagnostic statistic becomes

$$\widetilde{W}_i^M = D_i^M(\widehat{M}_{M(i)}\widehat{Q}_{M(i)}^{-1}\widehat{M}_{M(i)};\ (n-1)\hat{\sigma}_M^2) = \frac{(EIC_i^M)^T \cdot \widehat{M}_{M(i)}\widehat{Q}_{M(i)}^{-1}\widehat{M}_{M(i)} \cdot EIC_i^M}{(n-1)\hat{\sigma}_M^2} .$$

Substituting formula (3.19) in above formula, the diagnostic statistic can be written as

$$\widetilde{W}_i^M = (n-1)\psi^2(\frac{r_{i(i)}^M}{\hat{\sigma}_M})\mathbf{x}_i^T\widehat{Q}_{M(i)}^{-1}\mathbf{x}_i \ . \tag{3.35}$$

For GM-estimate case, the corresponding diagnostic statistic is

$$\widetilde{W}_i^{GM} = (n-1)\eta^2(\mathbf{x}_i, \frac{r_{i(i)}^{GM}}{\hat{\sigma}_{GM}})\mathbf{x}_i^T\widehat{Q}_{GM(i)}^{-1}\mathbf{x}_i \ . \tag{3.36}$$

3.2.7 Summary of Diagnostic Statistics

So far, we have discussed several diagnostic statistics. We can group them based on residuals, leverages, distances and change in fit.

(1). The diagnostic statistics based on residuals

standardized residuals : $r_{si} = \frac{r_i}{\hat{\sigma}\sqrt{1-h_i}}$

studentized residuals : $r_i^* = \frac{r_i}{\hat{\sigma}_{(i)}\sqrt{1-h_i}}$

robustly standardized residuals : $r_{si}^{\phi} = \frac{r_i^{\phi}}{\hat{\sigma}_{\phi}\cdot\sqrt{1-w_i^{\phi}}}$

robust deleted residuals : $r_{d(i)}^{\phi} = \frac{\mathbf{y}_i - \mathbf{x}_i^T\hat{\beta}_{\phi(i)}}{\hat{\sigma}_{\phi}}$

(for M-estimate) $r_{d(i)}^M \approx \frac{r_i^M}{\hat{\sigma}_M}[1 + \frac{wt_i^M\mathbf{x}_i^T\widehat{M}_M^{-1}\mathbf{x}_i}{1-w_i^M}]$

(for GM-estimate) $r_{d(i)}^{GM} \approx \frac{r_i^{GM}}{\hat{\sigma}_{GM}}[1 + \frac{wt_i^{GM}\mathbf{x}_i^T\widehat{M}_{GM}^{-1}\mathbf{x}_i}{1-w_i^{GM}}]$

(2). The diagnostic statistics based on leverages

leverages : $h_i = \mathbf{x}_i^T (X^T X)^{-1} \mathbf{x}_i$

robust leverages : $w_i^{\phi} = tr\{[-\frac{\partial}{\partial \hat{\beta}_{\phi}} \phi(z_i, \hat{\beta}_{\phi}, \hat{\sigma}_{\phi})] \hat{\sigma}_{\phi} \widehat{M}_{\phi}^{-1}\}$

(for M-estimate) $w_i^M = \psi'(\frac{r_i^M}{\hat{\sigma}_M}) \mathbf{x}_i^T \widehat{M}_M^{-1} \mathbf{x}_i$

(for GM-estimate) $w_i^{GM} = \eta'(\mathbf{x}_i, \frac{r_i^{GM}}{\hat{\sigma}_{GM}}) \mathbf{x}_i^T \widehat{M}_{GM}^{-1} \mathbf{x}_i$

robust Mahalanobis distance : $\mathbf{x}_i^T \widehat{M}_{\phi}^{-1} \mathbf{x}_i$

Note:

The robust leverages are the general form of the leverage obtained by the design matrix X. Since some values of $\psi'(\frac{r_i^M}{\hat{\sigma}_M})$ and $\eta'(\mathbf{x}_i, \frac{r_i^{GM}}{\hat{\sigma}_{GM}})$ may be zero for large residuals, the robust Mahalanobis distance can be used as an alternative statistic to measure the influence of the observations. Also we can see that if the ordinary least squares is applied, the robust Mahalanobis distances have the same values as the leverages.

(3). The diagnostic statistics based on change in fit

Welsch-Kuh distance (DFFITS) : $DFFITS_i = |r_i^*| \sqrt{\frac{h_i}{1-h_i}}$

standardized change in fit : $SCF_i^{\phi} = |\frac{\hat{y}_i^{\phi} - \hat{y}_{(i)}^{\phi}}{\hat{\sigma}_{\phi}}|$

(for M-estimate) : $SCF_i^M = wt_i^M \cdot |r_{si}^M| \cdot \frac{\mathbf{x}_i^T \widehat{M}_M^{-1} \mathbf{x}_i}{\sqrt{1-w_i^M}}$

(for GM-estimate) : $SCF_i^{GM} = wt_i^{GM} \cdot |r_{si}^{GM}| \cdot \frac{\mathbf{x}_i^T \widehat{M}_{GM}^{-1} \mathbf{x}_i}{\sqrt{1-w_i^{GM}}}$

(4). The diagnostic statistics based on distances

Cook's distance : $C_i = \frac{r_{Ai}^2}{p} \frac{h_i}{(1-h_i)}$

robust Cook's distance : $C_i^{\phi} = \frac{(SIC_i^{\phi})^T \cdot \widehat{M}_{\phi} \cdot SIC_i^{\phi}}{(n-1)^2 p \hat{\sigma}_{\phi}^2}$

(for M-estimate) : $C_i^M \approx \frac{\psi^2(r_i^M/\hat{\sigma}_M)}{p} \frac{\mathbf{x}_i^T \widehat{M}_M^{-1} \mathbf{x}_i}{(1-w_i^M)^2}$

(for GM-estimate) : $C_i^{GM} \approx \frac{\eta^2(\mathbf{x}_i, r_i^{GM}/\hat{\sigma}_{GM})}{p} \frac{\mathbf{x}_i^T \widehat{M}_{GM}^{-1} \mathbf{x}_i}{(1-w_i^{GM})^2}$

Welsch's distance : $W_i^2 = (n-1) r_i^{*2} \frac{h_i}{(1-h_i)^2}$

robust Welsch's distance : $W_i^{\phi 2} = \frac{(EIC_i^{\phi})^T \widehat{M}_{\phi(i)} EIC_i^{\phi}}{(n-1)\hat{\sigma}_{\phi}^2}$

(for M-estimate) $W_i^{M2} = (n-1)\psi^2(\frac{r_{i(i)}^M}{\hat{\sigma}_M}) \mathbf{x}_i^T \widehat{M}_{M(i)}^{-1} \mathbf{x}_i$

(for GM-estimate) $W_i^{GM2} = (n-1)\eta^2(\mathbf{x}_i, \frac{r_{i(i)}^{GM}}{\hat{\sigma}_{GM}}) \mathbf{x}_i^T \widehat{M}_{GM(i)}^{-1} \mathbf{x}_i$

4 Case Study

In this section, we apply several diagnostic statistics which were discussed in previous part to two data sets. These two data sets have been well explored in the literature.

4.1 The Land Use and Water Quality Data

The first data set we used is the Land use and water quality data which was collected by Haith (1976). Allen and Cady (1982) gave the details of this data, Simpson, Ruppert and Carroll (1992) gave some diagnostic results. In this data set, there are five variables :

N = total nitrogen; mean concentration, in milligrams per liter, based on samples taken at regular intervals during the spring, summer, and fall months,

AC = active agriculture; percentage of land area currently in agricultural use (cropland, pasture, etc),

FR = forest; percentage of land area in forest, forest brushland, and plantation,

RD = residential; percentage of land area in residential use

CI = commercial and/or industrial; percentage of land area in either commercial or manufacturing use.

4.1.1 The Model and The Estimates

There are 20 river basins (New York area) in which data had been collected. The total nitrogen content is treated as a measure of water quality. To measure the quantitative relationship between water quality and land use, initially, the following linear model was suggested

$$N_i = \beta_1 + \beta_2 AC + \beta_3 FR + \beta_4 RD + \beta_5 CI + \varepsilon_i \ .$$

After checking the data, it is suggested that between variables RD and CI, there is a strong positive association ($\rho = 0.86$). To alleviate the collinearity, these two variables are summed and form a new variable $UR = RD + CI$ (See Simpson et al (1992)). The linear model becomes

$$N_i = \beta_1 + \beta_2 AC + \beta_3 FR + \beta_4 UR + \varepsilon_i \ .$$

Based on the given data, the ordinary least squares estimate and several robust estimates (Huber's M-estimate, Mallows' GM-estimate and Schweppe's GM-estimate) are conducted. We use Huber's ψ function with $k = 1$ for Huber's M-estimate. For the Mallows' and Schweppe's GM-estimates, we used same functional forms as as formulas (2.7) and (2.8). The functions $w(x)$ and $v(x)$ are defined the same way as in part two. (*i.e.* $w(x) = \sqrt{1 - h_i}$ and $v(x) = 1/w(x)$). Based on above setting, we can get estimates for this data set. The estimates results are shown on Table 9. The values in parentheses are the standard deviation of the estimators.

	OLS	M(Huber)	GM(Mallows)	GM(Schweppe)
intercept	1.428598	2.423382	2.667217	2.700284
	(1.292693)	(0.917729)	(0.663689)	(0.576006)
AC	0.008503	−0.002815	−0.008152	−0.008371
	(0.01581560)	(0.010264)	(0.008018)	(0.007102)
FR	−0.008449	−0.019797	−0.023581	−0.024064
	(0.01446809)	(0.010486)	(0.008085)	(0.007073)
UR	0.029401	0.005459	0.087582	0.088769
	(0.027639)	(0.018866)	(0.065829)	(0.057825)

Table 9: The estimators of Land data and their variance estimation

4.1.2 Outliers and Influence Point Detection

(1). The normal probability plot for scaled residuals

The Normal probability plots of OLS studentized residuals and three types of scaled robust residuals are plotted (see Figure 3). (The scaled robust residual is defined as $r_i^{\phi}/\hat{\sigma}_{\phi}$). From the OLS studentized residuals, we can see that two points have large residuals. They are case 5 (the value is −5.582) and case 7 (the value is 4.061). The Huber's M-estimate shows cases 7 and 19 have large values of scaled residuals (the values are 5.167 and −3.805 respectively). Both of Mallows' and Schweppe's type scaled robust residuals suggest that case 5 has extreme residual (the values are −24.976 and −27.058 respectively). In both of those estimates, we also can find case 19 has large value, with values −4.986 for Mallows' type GM-estimate and −5.301

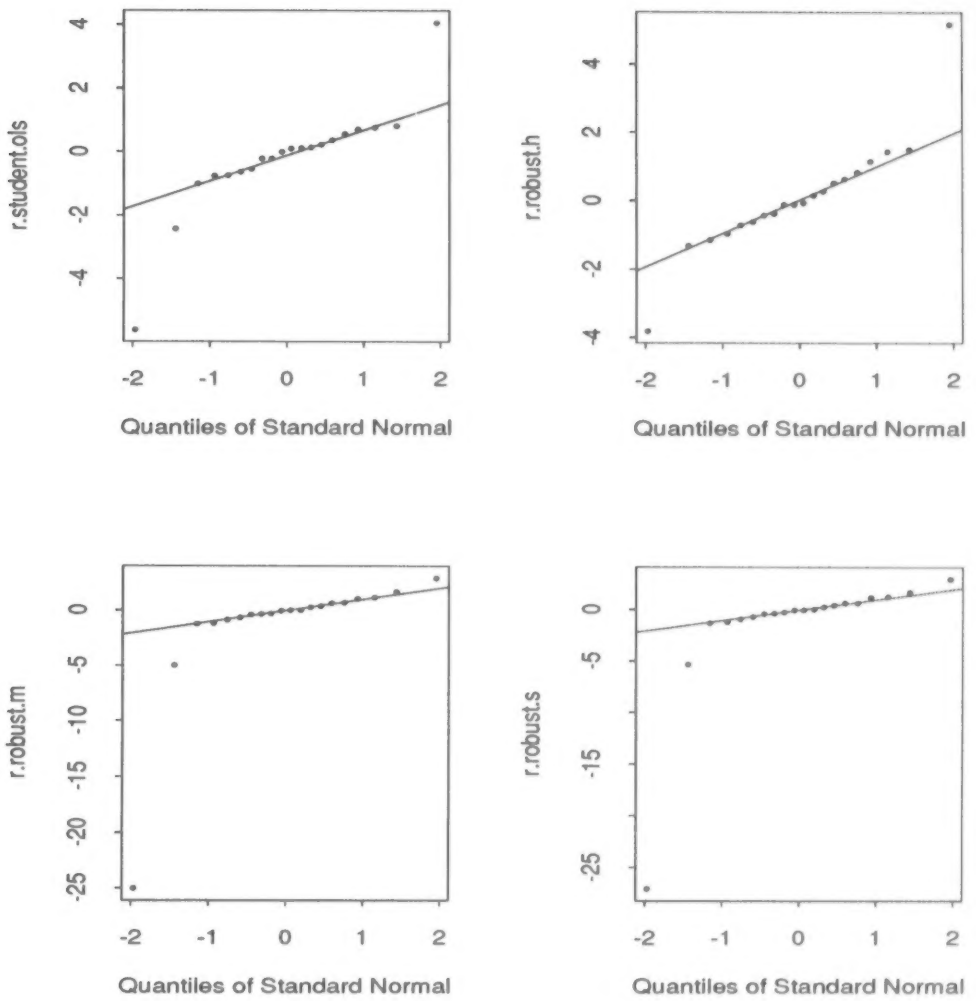


Figure 3: Normal probability plot for residuals ols: ordinary least squares, h: Huber's M-estimate, m: Mallows' GM-estimate, s: Schweppe's GM-estimate

for Schweppe's type GM-estimate. But if we compare those values with the values for case 5, we can see they are relatively small.

After checking the data set, we found that those cases which have large values on the residuals are Hackensack (case 5), Oswegatchie (case 19) and Fishkill (case 7).

(2). Leverages, robust leverages and robust Mahalanobis distance

For the leverages, the cutoff is 0.4 ($2p/n$). After checking the results, we found that case 5 has an extremely large leverage value ($h_5 = 0.957$). Cases 3, 18, and 19 have their values close to 0.4 (they have values 0.366, 0.335 and 0.315 respectively). For the robust leverages, Huber's M-estimate shows that case 5 has value 0.999, case 3 and 18 have values exceed 0.4 (the values are 0.530 and 0.498 respectively). For Mallows GH-estimate, we get the large values are cases 3, 4, 6 and 18 (they have values 0.537, 0.596 , 0.505 and 0.495 respectively). Similarly, for Schweppe's GM-estimate, we found same cases have large values (the values are 0.674, 0.654, 0.519 and 0.608 for case 3, 4, 6 and 18 respectively). From the data, case 5 is obviously an unusual point. The reason we obtained a small value on case 5 is that robust leverages are zero if the residuals are large. It indicates that robust leverage may not be a good statistic. An alternative statistic which can be applied in diagnosis is robust Mahalanobis distance.

For Mallows and Schweppe's GM-estimates, both of them have large values of robust Mahalanobis distance (71.387 for Mallows type GM-estimate and 66.704 for Schweppe's type GM-estimate). Those values are extremely large comparing with the values of other cases. It indicates case 5 is a possible influential point. The cases (3, 4, 6 and 18) which were detected by robust leverage are not likely to be influential points in the robust regressions.

After taking a look, we can see that those cases which have large leverage values and large robust Mahalanobis distance are Hackensack (case 5), Oatka (case 3), Raquette (case 18), Owegatchie (case 19).

(3). Welsch-Kuh distance (DFFITS) and Standardized Change of fit (SCF)

By Belsley, Kuh and Welsch's suggestion, we choose $2\sqrt{p/n}$ as our cutoff which equals 0.894. From the calculation result of OLS case, we can see that the $DFFITS$ values of cases 5, 19 and 7 exceed the criterion value (they are 26.216 , 1.648 and 1.057 respectively). For the Standardized Change of fit (SCF) of robust estimate, we can see that cases 5 and 3 have large values (511.787 and 0.934) for Huber's M-estimate. (Considering the value of case 3, although it exceeds 0.894, the reasonable conclusion is that we only think case 5 is an extreme case). If we check the SCF for the GM-estimate cases, we can find that cases 5 and 7 have large values. For Mallows' GM-estimate, the SCF values are 14.867 and 1.963 respectively. For Schweppe's GM-estimate, the values are 13.892 and 1.835. (For case 5 and case 7,

since the values for case 7 are relatively small in comparing with the values of case 5, a reasonable conclusion is that the case 5 is a possible influential point).

After checking, we can find that those cases which have large values of DFFITS and SCF are Hackensack (case 5), Owegatchie (case 19) and Fishkill (case 7) .

(4). Cook's distance

The criterion value for Cook's distance is $F(0.5, 4, 16)$ which equals 0.8758. The computation results show that case 5 has a extremely value in OLS and Huber's M-estimate. Their values are 59.557 and 65536.5 respectively. The Cook's distance for GM-estimates gives a similar result that case 5 has bigger value compare with other cases, but it tends to be relatively small value in comparing with the values obtained by OLS or M-estimate case. (0.774 for Mallows' GM and 0.723 for Schweppe's GM). From Cook's distance, we can see that case 5 (Hackensack) is an obviously influential point in the data set.

(5). Welsch's distance

The criterion value for Welsch's distance is $3\sqrt{p}$, it equals 6 in this data set. We found that case 5 has an extreme value (548.704 for OLS case, 20.850 for Huber's M-estimate, 35.990 for Mallows' GM-estimate and 23.554 for Shweppe's GM-estimate). It also shows that case 19 has Welsch's distance greater than 6 for OLS case. But for the robust estimate, case 19 does not have large values. In Schweppe's

GM-estimate case, it also suggests that the case 7 requires us to pay attention (which has value 9.835). The modified Welsch's distance for robust estimates gave the similar results for case 5. (32.159 for Huber's M-estimate, 22.469 for Mallows' GM-estimate and 23.429 for Schweppe's GM-estimate).

4.1.3 Conclusion of the Diagnosis

From above results, we can see that case 5 is an influential point. If check Table 9, we can see that the estimators of OLS and Huber's M-estimate, Mallows' (or Schweppe's) GM-estimate are quite different. Basically, this is caused by unusual point case 5. We also conclude that case 7 (Fishkill) and case 19 (Owegatchie) are possible outliers in this data set.

	OLS	M	GM(Mallows)	GM(Shweppe)
intercept	1.698556	2.625213	2.663357	2.715427
	(0.762612)	(0.630076)	(0.608121)	(0.542684)
AC	0.002077	−0.008611	−0.009037	−0.009516
	(0.009382)	(0.006784)	(0.006529)	(0.005967)
FR	−0.013762	−0.023870	−0.024196	-0.024856
	(0.008571)	(0.007239)	(0.007003)	(0.006340)
UR	0.156235	0.132722	0.127248	0.126723
	(0.027948)	(0.020670)	(0.021538)	(0.036763)

Table 10: The estimators of Land data (case 5 is excluded) and their variance estimation

Since in this data set case 5 is such a severe design outlier, we redo the regression based on deleting case 5. Using the same model, we have the estimates of parameters show on Table 10.

Note: From this data set, we can have a clear look about the weakness of ordinary M-estimate which is only robust against outliers of response $\mathbf{y}$ but not robust against outliers in design point $\mathbf{x}_i$ as we mentioned several times before. This data has case 5 which is an extreme point in the design matrix X. The M-estimate does not give us a satisfactory estimator. On the other hand, the two types GM-estimates gave us robust estimates which fit the majority of the data. After case 5 has been deleted, the estimated parameters values are much more close for the different types of estimates.

4.2 The Hertzsprung-Russell Star Data

Now we use another data set to conduct the diagnostic study. The data set is Hertzsprung-Russell stars data which was discussed by Rousseeuw and Leroy (1987) and Hadi and Simonoff (1993).

This data set consists of 47 measurements of the logarithm of effective temperature at the surface of a star and the logarithm of the light intensity of the stars. Most of the cases have a direct relationship between the two variables, but there are four red giants (case 11, 20, 30, 34) have low temperature with high light intensity. From Figure 4, we can see that those cases are outliers.

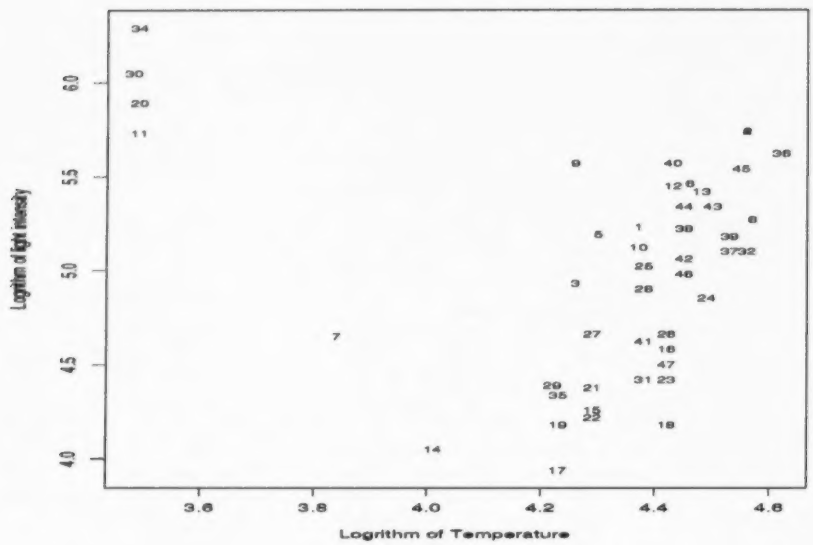


Figure 4: Scatterplot of Hertzsproung-Russell data

For this data set, the robust diagnostic statistics and the classical diagnostic statistics are applied in detecting outliers and influential points. We found that it is hard to detect those outliers if we use standardized residuals and deleted residuals. Also the classical diagnostic statistics like DFFITS and Cook's distance show us that cases 11, 14, 20, 30 and 34 are possible outliers (See Figure 5). For DFFITS, it gave values 0.3651, 0.4388, 0.5226, 0.6907 and 0.93533 respectively. We can see that case 14 has larger DFFITS value than case 11. If we use DFFITS as our diagnostic statistic, it will cause misleading and obtain the not suitable conclusion.

Same as the classical Cook's distance, it also shows that cases 11, 14, 20, 30 and 34 are possible outliers. The value at case 14 is larger than the value at case 10.

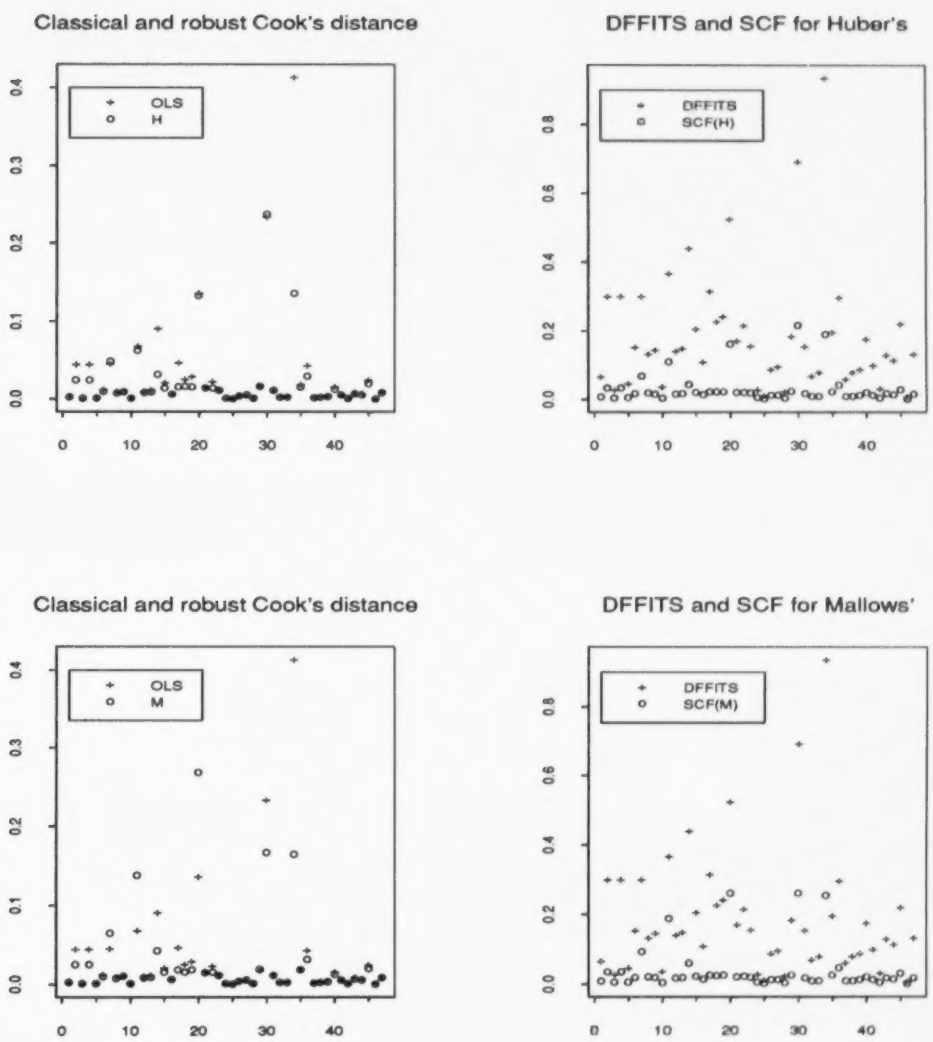


Figure 5: Plot of diagnostic statistics for Hertzsproung-Russell data

When we applied two robust diagnostic statistics, SCF (standardized change of fit which is an analogue to DFFITS) and robust Cook's distance (which is an analogue to classical Cook's distance), we found that those two statistics give us good results. Figure 5 shows that for Huber's M-estimate and Mallows' GM-estimate (from the computation results, Schweppe's GM-estimate has the similar results as well), the SCF and robust Cook's distance detect cases 10, 20, 30 and 40 are possible outliers or influential points.

When we consider the leverage, robust leverage and robust Mahalanobis distance, we can see that cases 11, 20, 30 and 34 have large leverage values as well as robust Mahalanobis distance. The robust leverage shows us cases 10 and 20 have large values but case 30 and 34 do not (they have zero values). This is due to that case 30 and 34 have large residuals. This result shows again that the robust leverage may not be a good diagnostic statistic. We can use robust Mahalanobis distance instead of using robust leverage, in analogy to the leverage in the classical case.

References

[1] Allen, D. M. and Cady, F. B. (1982) *Analyzing Experimental Data by Regression* Lifetime Learning Publications, Belmont, Califonia

[2] Belsley, D. A., Kuh, E., Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity* John Wiley & Sons, Inc., New York

[3] Chatterjee, S. & Hadi, A. S. (1986) Influential Observations, High Leverage Point, and Outliers in Linear Regression *Statistical Science* **Vol. 1, No. 3** 379-416

[4] Cook, R. D. (1977) Detection of Influential observations in Regression *Technometrics* **22** 495-508

[5] Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans* Society for Industrial and Applied Mathematics

[6] Hadi, A. S. and Simonoff, J. S. (1993) Procedures for the Identification of Multiple Outliers in Linear Models *Journal of the Amer. Statist. Assoc.* **88, No. 424** 1264-1272

[7] Haith, D. A. (1976) land Use and Water Quality in New York River *Journal of the Environmental Engineering Division, Proceedings of the American Society of Civil Engineers* **102** 1-15

[8] Hampel, F. R. (1974) The Influence Curve and Its Role in robust estimation. *Journal of Amer. Statist. Assoc.* **69** 383-393

[9] Hampel, F. R. et al (1986) *Robust Statistics : The Approach Based on Influence Functions* John Wiley & Sons, Inc., New York

[10] Hinkley, D. V. (1977) Jackknifing in Unbalanced Situations *Technometrics* **19** 285-292

[11] Huber, P. J. (1981) *Robust Statistics* John Wiley & Sons, Inc., New York

[12] Miller, R. G. (1974a) The Jackknife - a review *Biometrika* **61** 1-15

[13] Miller, R. G. (1974b) An Unbalanced Jackknife *Ann. Statist.* **2** 880-891

[14] Quenouille, M. H. (1949) Approximate Tests of Correlation in Time Series. *Journal of the Royal Statistical Society* B **11**, 68-84

[15] Quenouille, M. H. (1956) Notes on Bias in Estimation. *Biometrika* **43**, 353-360

[16] Rao, C. R. (1973) *Linear Statistical Inference and Its Applications* John Wily & Sons, Inc. New York

[17] Rousseeuw, P. J. and Leroy, A. M. (1987) *Robust Regression and Outlier Detection* John Wiley & Sons, Inc., New York

[18] Simpson, G., Ruppert, D., Carroll, R. J. (1992). On One-Step GM Estimates and Stability of Inferences in Linear Regression *Journal of the American Statistical Association* **Vol. 87, No. 418** 439-450

[19] Tukey, J. (1958) Bias and Confidence in not Quite Large Sample (abstract). *Ann. Math. Statist.* **29** 614

[20] Velleman, P. F. and Welsch, R. E. (1981) Efficient Computing of Regression Diagnostics *Amer. Statist.* **35** 234-242

[21] Wu, C. F. J. (1986) Jackknife, bootstrap and other resampling methods in regression analysis *Ann. Statist.* **14** 1261-1295